Aalto University
School of Science
Master's Programme in Computer, Communication and Information Sciences

Mihaela Mihaylova

A Study of LDL Aggregation and its Correlation with the LDL Lipidome and Clinical Data in Bariatric Surgery Patients

Master's Thesis
Espoo, August 14, 2019

Supervisors: Docent Lasse Leskelä, Aalto University
Advisor: Dr. Matti Pirinen, University of Helsinki

Aalto University
School of Science
Master's Programme in Computer, Communication and Information Sciences

ABSTRACT OF MASTER'S THESIS

Author: Mihaela Mihaylova

Title:
A Study of LDL Aggregation and its Correlation with the LDL Lipidome and Clinical Data in Bariatric Surgery Patients

Date: August 14, 2019 **Pages:** 88

Major: Data Science **Code:** SCI3095

Supervisors: Docent Lasse Leskelä

Advisor: Dr. Matti Pirinen

Atherosclerotic cardiovascular disease (ASCVD), also known as coronary artery disease (CAD), is one of the leading causes of death in the world.[22] A consensus has been reached that the main cause of ASCVD are low-density lipoproteins (LDL).[16] ASCVD develops in the innermost layer of the coronary artery wall (intima). Once LDL particles enter the wall, they are retained, modified, and accumulate there.[35] There are several well-known risk factors of ASCVD, among which obesity, smoking, hypertension and LDL cholesterol concentration in the plasma.[35] A novel approach to assessing the risk of ASCVD however suggests that not only the concentration, but also the susceptibility of LDL particles to aggregate plays a role in ASCVD. [35] This makes investigating LDL aggregation particularly important.

This thesis studies LDL aggregation of one particular cohort - obese people who underwent bariatric surgery, a standard weight-loss procedure. The goal of the thesis is to investigate whether bariatric surgery affects LDL aggregation, the structure of LDL, and clinical parameters of the patients. In addition, it aims to establish whether LDL aggregation in bariatric surgery patients is correlated with the structure of LDL and selected clinical parameters.

The thesis is thus built upon four main points. First, it focuses on creating a nonlinear mixed-effects model of LDL aggregation and obtaining a single quantitative measure for it. Second, it investigates whether there is a significant difference in LDL aggregation in patients before and after bariatric surgery. Third, it studies correlations of LDL aggregation and the core and surface lipids, contained in LDL, known as the *LDL lipidome*, as well as clinical data of the patients. Fourth, it investigates whether there is a significant difference in the LDL lipidome lipids and clinical parameters in the patients before and after the operation.

The results indicate that bariatric surgery does not appear to affect LDL aggregation. However, it appears to affect the surface structure of LDL. In addition, patients exhibit improvement of blood pressure, glucose levels, as well as BMI, after the operation.

Keywords: LDL, aggregation, lipid, atherosclerosis, bayesian, mixed-effects, nonlinear, model, correlation, bariatric, surgery

Language: English

Acknowledgements

First, I would like to thank the lead of the Atherosclerosis group at Wihuri Research Institute, Doc. Kati Öörni for the opportunity to be part of this project, as well as her patience and ability to explain complex biochemistry to a statistician. This project sparked my curiosity and helped me discover my passion for biostatistics. I would also like to thank my supervisor Doc. Lasse Leskelä and my advisor Matti Pirinen for their patience and help. I would also like to thank the whole Stan community for their tireless dedication to promote Bayesian statistics and their readiness to be helpful.
Last but not least, I would like to thank my friends and family for helping me keep my chin up and making me laugh, even when project work was not going smoothly.

Espoo, August 14, 2019

Mihaela Mihaylova

Abbreviations and Symbols

Abbreviations

LDL	Low-density lipoprotein
ASCVD	Atherosclerotic Cardiovascular Disease
CAD	Coronary Artery Disease
DLS	Dynamic Light Scattering
FDR	False Discovery Rate
FWER	Familywise Error Rate
LOO-CV	Leave-One-Out Cross-Validation
ELPD	Expected Log-Pointwise Predictive Density
CLT	Central Limit Theorem
PPC	Posterior Predictive Check
MAE	Mean Absolute Error

Symbols

N(-,-)	Normal distribution
Gamma(-,-)	Gamma distribution
$\mathbf{X}$	set of explanatory variables
x	explanatory variable
$\mathbf{y}$	observed response
p(x)	probability distribution function
p(x,y)	joint probability distribution function
p(x|y)	conditional probability
E	expected value
θ	model parameters
$\hat{\theta}$	estimator for model parameters

μ	mean
σ	standard deviation
g(.)	link function
η	linear predictor

Contents

Chapter 1

Introduction

Cardiovascular diseases are the major cause of death in developed and an increasing cause of death in developing countries. [18] Thus their cause, treatment, and prevention have become a focus of research for many. The role of LDL in ASCVD, one of the most spread cardiovascular diseases, has been of particular interest. Based on a large body of evidence from clinical, epidemiological, and genetic studies, an agreement has been reached among researchers that LDL is the main cause of ASCVD. [16]
LDL particles can efficiently cross the innermost layer of the coronary artery, the coronary intima. They accumulate at the sites of the intima, which are prone to plaque formation, and are turned into plaque. [16] If high concentration of LDL in the plasma is maintained for a long period of time, the amount of plaque increases, resulting in ASCVD, characterized by a decrease in blood flow to the heart muscle, and the arterial walls getting more rigid and prone to rupture. If the plaque in the arterial wall ruptures, the platelets start clumping and form a blood clot, which could potentially lead to complete blocking of the artery, which means that part of the heart is deprived of oxygen, resulting in a heart attack.
Recent studies, however, suggest that not only the concentration of LDL in the plasma, but also the quality of LDL might play a role in CAD. And, more precisely, LDL's susceptibility to aggregate may be used as a predictor of future CAD deaths. [35] One group of people who are particularly at risk of ASCVD are obese people. Obesity is considered a major risk factor for ASCVD, as it is associated with other major risk factors, leading to ASCVD, such as hypertension, hyperglycemia, and hypercholesterolemia. [25]

This thesis focuses on studying the LDL aggregation of patients who underwent bariatric surgery —an operation which helps obese people lose weight, by making changes to their digestive system. [32] Available before- and after-operation data of the patients is used. There are three main types

of data considered —LDL aggregation data, LDL lipidome data —the types of lipids found in the surface and core of LDL, as well as patients' clinical data. These are split into four datasets —an LDL aggregation dataset, an LDL surface lipids dataset, an LDL core lipids dataset, as well as a set of clinical data.

The goal of the thesis is to investigate whether bariatric surgery affects LDL aggregation, the structure of LDL, and clinical parameters of the patients. In addition, it aims to establish whether LDL aggregation in bariatric surgery patients is correlated with the structure of LDL and selected clinical parameters. The exact questions which this thesis aims to answer are given below:

- Is there a significant difference in LDL aggregation of bariatric patients before and after the operation?

- Are there any significant changes in the structural lipids of LDL particles of the bariatric patients before and after the operation?

- Are there any significant correlations between LDL aggregation and LDL structural lipids before and after the operation?

- Are there any significant changes in the clinical parameters of the bariatric patients before and after the operation?

- Are there any significant correlations between LDL aggregation and clinical parameters before and after the operation?

- Are there significant correlations between change of LDL aggregation and change in the structural lipids and clinical parameters of the patients, as a result of the operation?

The steps taken towards achieving the goal of the thesis are described in the following chapters. In Chapter 2, the necessary background knowledge on LDL, LDL aggregation, and measuring LDL aggregation is provided. In Chapter 3, we look more closely at the work that has been done so far on the topic. And more precisely, we focus on the paper by Ruuth et al[35] which suggested using LDL aggregation as a potential diagnostic for ASCVD. The similarities and differences of the modeling approaches of LDL aggregation in this thesis and the original paper are described. In Chapter 4, the datasets available for our study are described. In addition, the chapter describes the necessary data preparation done, which allows us to get a clearer idea of the

nature of the data and thus helps us identify the best steps to take, in order to answer the posed research questions. Furthermore, this chapter provides a brief review of the basics of Bayesian inference, a description of nonlinear mixed-effects models, as well as the mathematical formulation of the LDL aggregation model. We also discuss methods for model checking and model comparison, as well the statistical tests we use to obtain answers of the research quesitons.
Chapter 5 provides results of the tests conducted on the available data. Chapter 6 then provides interpretation of the results obtained and their implications. It also discusses some limitations of the work done in the thesis and ways to potentially improve it in following studies. Chapter 7 concludes the thesis by providing a brief summary of the work done.

Chapter 2

Background

The goal of this chapter is to provide the necessary background in biology, in order for the reader to be able to understand the research problem at hand, as well as some data preparation and modelling decisions. We first briefly discuss the nature, structure, and purpose of lipoproteins in the body, as well as the types of lipoproteins that can be found in humans. We then introduce the low-density lipoprotein (LDL), which is at the core of the research problem in this thesis. We focus on the structure of LDL and its role in the pathophysiology of atherosclerosis. In addition, the steps taken to initiate and measure LDL aggregation are discussed. We then end the chapter with a brief introduction to Dynamic Light Scattering (DLS) - the method used for measuring LDL aggregation, as this gives some insights in the data cleaning process described later on.

2.1 Lipoproteins

Lipoproteins are complex particles which consist of lipids, which are insoluble in water, and proteins, whose purpose is to transport these lipids in the body.

2.1.1 Lipids

Lipids (also known as fats) are a diverse group of organic molecules, which are poorly soluble in water. Lipids can be polar amphipathic or nonpolar (completely insoluble in water). Polar lipids are poorly soluble in water because they are amphipathic, which means that they have both hydrophilic and hydrophobic regions in the same molecule. Some of the more important nonpolar lipids are cholesterol and triglycerides. [7] Triglycerides are important for the body as they serve as a source of energy.[37] Cholesterol

is a crucial component of cell membranes and is used for making molecules, which are essential to the body, such as hormones, fat-soluble vitamins, and bile acids which help in digestion.[1]

2.1.2 Proteins

Proteins are macromolecules, which consist of long chains of smaller units attached to one another, called amino acids. Proteins perform a wide range of functions in the human body, such as antibodies, structural components, transport and storage, and others. [8]

2.1.3 Structure of Lipoproteins

Since essential lipids such as cholesterol and triglycerides are insoluble in water, they need to be transported within the body in lipoproteins. Lipoproteins are compound particles, containing both lipids and proteins whose main purpose is the transport of lipids in the body. [14]

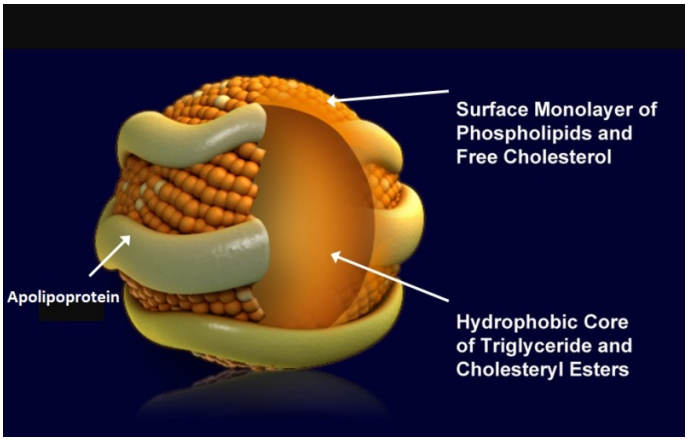

Figure 2.1: Structure of a lipoprotein, Source: [15]

Lipoproteins have a hydrophobic core which contains cholesterol esters and triglycerides. Their surface structure consists of free cholesterol, phospholipids, and apolipoproteins, which facilitate the formation and function of lipoproteins. [15]

2.1.4 Lipoprotein Classes

Plasma lipoproteins are divided into seven classes, based on their size, lipid composition, and types of apolipoproteins. [15] The seven classes are, as

follows: chylomicrons, chylomicron remnants, very-low-density lipoproteins (VLDL), intermediate-density lipoproteins (IDL), low-density lipoproteins (LDL), high-density lipoproteins (HDL), Lipoprotein-a (Lip-a).

These perform different functions in the body, however, for the sake of brevity, we would not go into more detail. The figure below shows the types of lipoproteins with their size ranges.

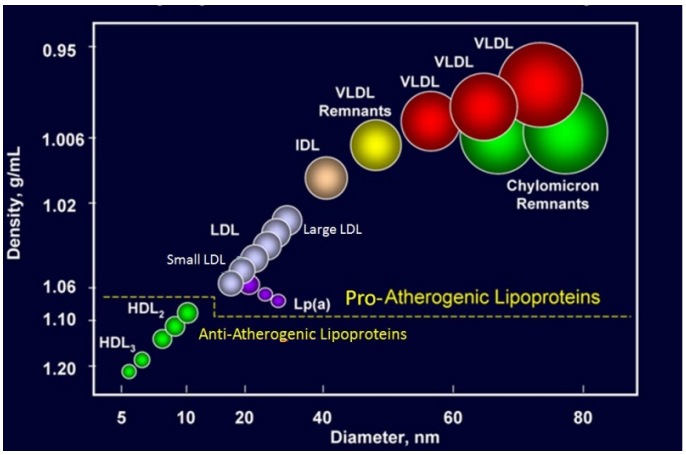

Figure 2.2: Classes of lipoproteins, Source: [15]

The majority of lipoproteins in the plasma, including LDL, which is in the focus of this thesis, are pro-atherogenic. HDL is the only anti-atherogenic plasma lipoprotein. The pro-atherogenic lipoproteins promote the development of atherosclerosis - the underlying condition for ASCVD. We now focus on the structure and function of LDL, and the way it contributes to atherosclerosis.

2.2 LDL

Low-density lipoproteins, or LDL, are derived from VLDL and IDL, and have very high content of cholesterol. LDL is, in fact, the main carrier of cholesterol in the circulation.[15] An LDL particle consists of one molecule of Apo B-100 protein, triacylglycerol (also known as triglyceride) and esterified cholesterol in the core, and phospholipid monolayer and unesterified cholesterol forming the surface.[15] A figure showing the structure of an LDL particle is given below.

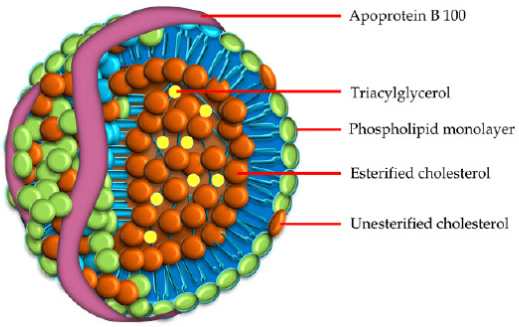

Figure 2.3: Structure of an LDL lipoprotein, Source: [11]

LDL particles vary in size and density. High numbers of small LDL particles present in the plasma are usually associated with obesity and inflammatory states. In addition, small LDL particles are supposed to be more pro-atherogenic. [15]

2.3 Role of LDL in Atherosclerosis

2.3.1 What is Atherosclerosis?

Atherosclerosis is the underlying reason for ASCVD.[22] It is a condition in which plaque builds up inside the arteries. The plaque consists of fat, cholesterol, calcium and other substances found in the blood. As the plaque buildup expands, it becomes more difficult for oxygen-rich blood to reach the heart and other body parts. This can result in heart attack, stroke, or, potentially, death. [31] For this reason, atherosclerosis has been a subject of interest of many researchers in the past decades. Based on evidence from a large amount of various clinical, epidemiological, and genetic studies, consensus has been reached among researchers that LDL is the main cause of atherosclerosis. [16]

2.3.2 How Does LDL Contribute to Atherosclerosis?

An artery wall consists of two main layers - intima and media. The inner layer, the intima, consists of a single layer of endothelial cells - the endothelium, and internal elastic lamina. [22]

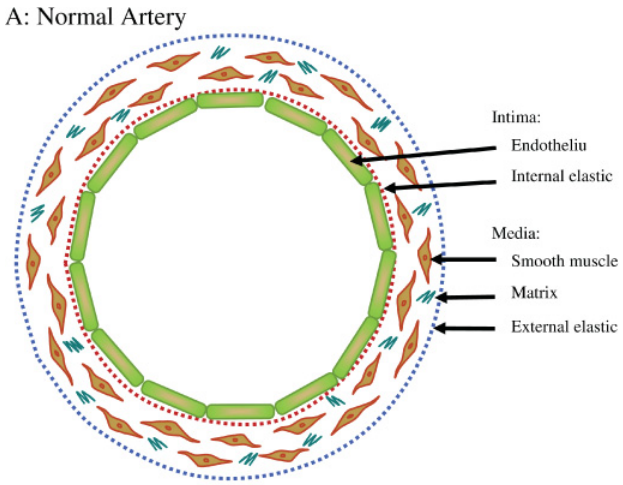

Figure 2.4: Structure of a normal artery, Source: [22]

The endothelium is prone to activation, due to inflammation, and damage caused by risk factors, such as high blood pressure, smoking, and obesity.[22] Once activated or damaged, the endothelium allows the entry of LDL and monocytes into the artery wall.[22] Once entering the artery, monocytes dif-

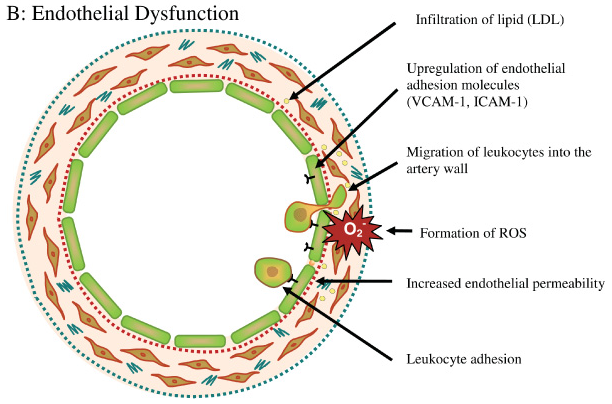

Figure 2.5: An artery wall with a damaged endothelium, Source: [22]

ferentiate into macrophages which take up the LDL particles and become foam cell macrophages. As a result, lesions, known as "fatty streaks" are formed. This is considered to be the onset of atherosclerosis.[22]

Fatty streaks may be precursors of larger atherosclerotic plaques, but they

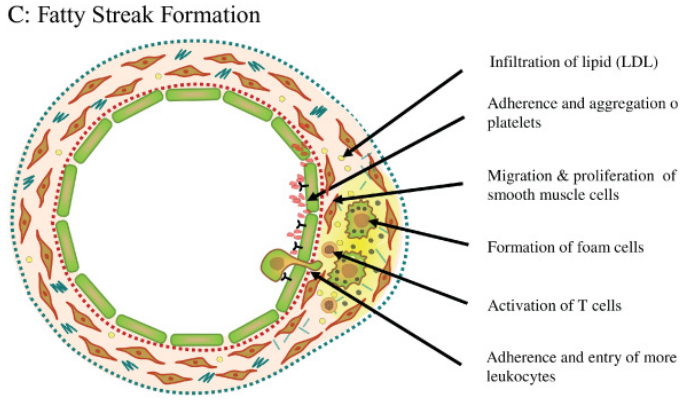

Figure 2.6: Fatty Streak Formation, Source: [22]

may also regress. A progression of fatty streaks is caused by the formation of a necrotic core and a fibrous cap. [22]

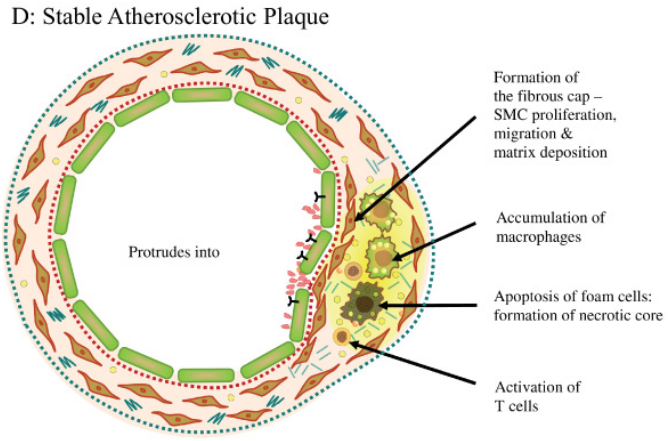

Figure 2.7: Formation of Stable Atherosclerotic Plaque, Source: [22]

Foam cells macrophages, full of LDL, start dying and release their contents, which helps in the formation of a necrotic core. The released content of the cytoplasm of the foam cells leads to the accumulation of extracellular lipids which induce inflammation. In addition, vascular smooth muscle cells migrate from the media layer into the intima and form a fibrous cap around the necrotic core. [22] The plaque could then grow and change. A plaque which has a large necrotic core, high content of inflammatory cells,and thin

fibrous cap is unstable, and is very prone to rupture. [22]

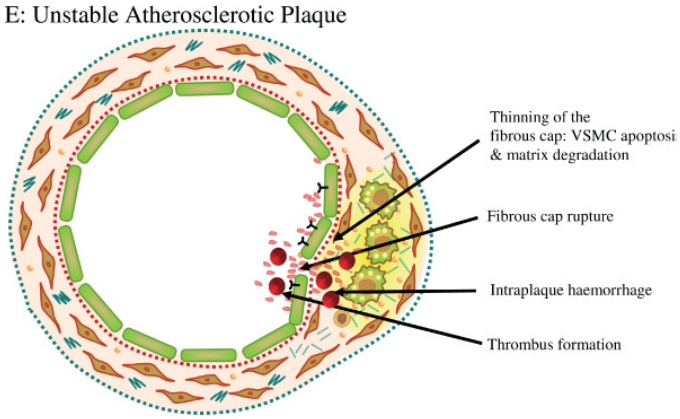

Figure 2.8: Unstable Atherosclerotic Plaque, Source: [22]

If the unstable plaque ruptures, platelets in the blood start forming a blood clot (thrombus). The supply of oxygen-rich blood to the heart is thus cut off, resulting in a heart attack and, potentially, death. [22]

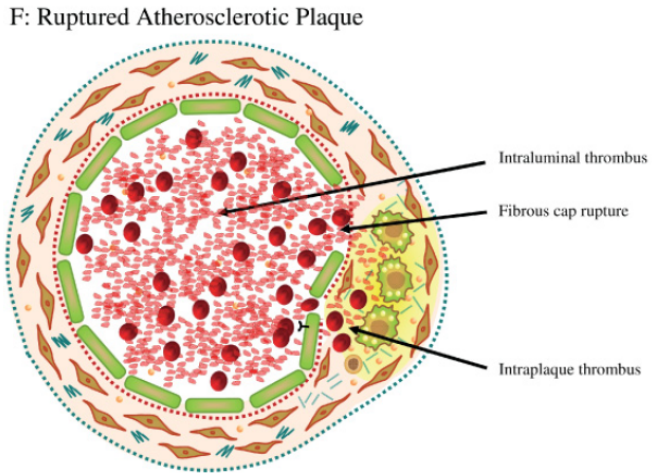

Figure 2.9: A Ruptured Unstable Plaque Causes Thrombus, Source: [22]

The rupture of the plaque is sudden, which makes it very dangerous, as the reaction time for treating the condition is fairly short. Thus, developing methods to assess a person's risk of atherosclerotic death is important, as it allows taking steps towards its prevention.

2.3.3 LDL as a Predictor of Cardiovascular Atherosclerotic Deaths

As mentioned earlier, high concentration of LDL is the main factor for atherosclerosis, however, the quality of LDL is also thought to be a contributor.[35] In this context, the study by Ruuth et al [35] defines the quality of LDL as its susceptibility to aggregate in the presence of the human recombinant secretory sphingomyelinase enzyme. In the study by Ruuth et al, the susceptibility of LDL to aggregate in the presence of the enzyme was predictive of future atherosclerotic deaths in patients with ASCVD, independently of traditional risk factors. [35]

2.4 Measuring LDL Aggregation

Having discussed the importance of LDL in predicting atherosclerotic deaths, it is now good to have a look at how LDL aggregation is measured, as this would allow us to better understand our data and be able to identify any errors in the dataset.
The process of measuring LDL aggregation consists of several steps:

1. Obtaining blood samples from the subjects of interest

2. Separating the blood cells from the plasma using centrifugation

3. Separating the LDL particles from the plasma using ultracentrifugation [35]

4. Human recombinant secretory sphingomyelinase enzyme was added to the isolated plasma LDL [35]

5. Once the LDL particles start aggregating, the radius of the aggregates is measured throughout time, using Dynamic Light Scattering (DLS).

It is good for the reader to get a brief overview of how Dynamic Light Scattering works, as it would put into context some of the decisions we make later on when dealing with problematic data points. In addition, it could potentially be helpful to any future reader who is on the statistics side of a biochemistry-related project making use of DLS.

2.4.1 Dynamic Light Scattering

In a solution, macromolecules are hit by the molecules of the solvent, leading to random motion, known as Brownian motion.[44]The moving macromolecules scatter light, with their motion adding randomness to the phase of the scattered light, so that when the light from several particles is added together, it results in a destructive or constructive interference. This, in turn, leads to time-dependent intensity fluctuations of the scattered light. [44] In DLS, the time-dependent fluctuations of the scattered light are measured by shooting a laser through the solution and registering the diffusion of the macromolecule using a fast photon counter.[44] The registered fluctuations are directly related to the macromolecule's rate of diffusion through the solvent, which is, in turn, related to the hydrodynamic radius of the macromolecule. [44] So, in effect, what DLS measures is the hydrodynamic radius of the macromolecule - that is, the radius of perfectly spherical particle which diffuses light at the same speed as the macromolecule of interest. [33]

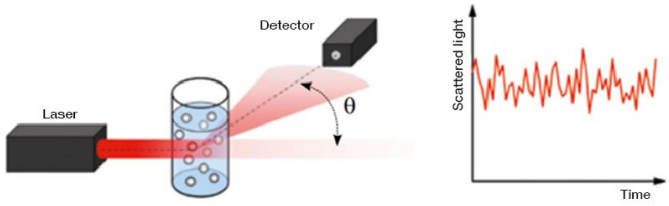

Figure 2.10: Basic Dynamic Light Scattering measurement system, Source: [33]

The measurements of scattered light in DLS are made over a short time period, allowing the instrument to monitor the particles' movement. As the particles move, the intensity of the scattered light fluctuates over time. Smaller particles move faster than large particles, thus resulting in faster fluctuations. The generated intensity trace is then used to generate an intensity autocorrelation function. [33] The autocorrelation function is essentially used to describe how long a particle stays at the same position in a solution.

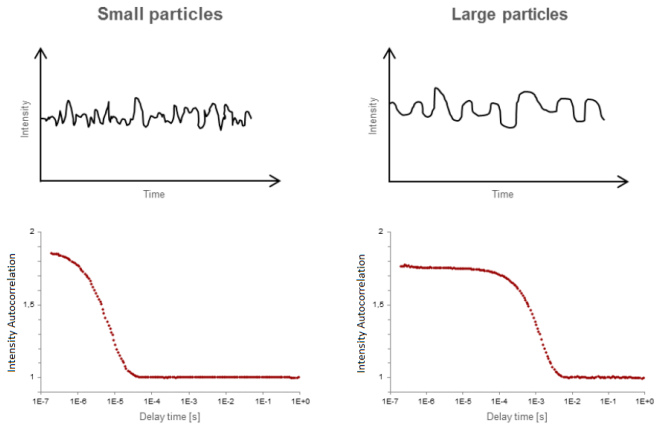

Figure 2.11: Comparison of Intensity and Correlation plots of small and large particles [33]

As one can see, the upper part of the autocorrelation graph is roughly constant, which indicates the particle being in the same position at time t + Δt, as it was at time point t. The exponential decay which follows means that it changes its position with time. When the particle gets to a position where there is no similarity with its initial position, the correlation function goes flat again. The bottom flat part of the correlation graph is called a baseline. [33]

The reason why this is important, is because it is one of the criteria for determining the quality of measurements. One can set a certain baseline and if this is not reached within a predetermined time frame, the measurement is discarded. The reasoning behind this is that if the correlation function is decreasing too slowly, this means that the molecule whose size we are trying to measure is too big, which can result in inaccurate measurements. [12]

Now that we have provided the necessary biological background to understand the research question, we now continue to discussing the paper by Ruuth et al, which served as a main inspiration for this thesis.

Chapter 3

Related Research

In this chapter we focus on the paper of Ruuth et al [35], which studies LDL aggregation and suggests it is predictive of future atherosclerotic deaths.

The study introduces the idea that not only the concentration, but also the quality of LDL contributes to ASCVD [35]. In this context, the "quality" of LDL is defined as its susceptibility to aggregate in the presence of the enzyme hrSMase. In the experiment described, LDL is isolated from the plasma of participants from several groups, including healthy volunteers, participants in a health examination survey, participants with CAD in a study about the risk factors and genetics of heart disease, as well as a dietary study. [35] Next, the enzyme hrSMase is added to the LDL samples, which causes LDL particles to start aggregating. The size of the aggregates is then measured in time, using DLS. The data is then used to fit a model to describe LDL aggregation in time.

In order to describe LDL aggregation with a single number, the authors use the size of LDL aggregates at the 2-hour time point. The reason for using this specific time point is because it was established to clearly separate subjects whose LDL aggregates slowly from subjects whose LDL aggregates very fast. The size at the 2-hour time point is also found to be strongly positively correlated with the inflection point of the curve fitted. In addition, LDL aggregation is also found to depend on the structure of the LDL particles. And, most importantly, the presence of aggregation-prone LDL is found to be associated with future CAD deaths, independently from other traditional risk factors. [35]

It is now worth drawing parallels, as well as establishing the differences between the Ruuth et al study and this thesis project. This thesis can be seen as a follow-up study of the this paper, with focus on one particular study population - bariatric surgery patients.

To start with, while the Ruuth et al study uses LDL aggregation data

from several different studies, this thesis is based only on data from bariatric surgery patients. Additionally, while the subjects from the original study can be categorized in several different groups, the data in this thesis is paired - before- and after-operation data of the same patients, which allows for studying changes in LDL aggregation, the LDL lipidome, and clinical parameters, as a potential result of the operation.

Additionally, the amount of LDL aggregation data in the original study is significantly larger than the amount of data used in this thesis, which makes it harder to develop a reliable, generalizable model for LDL aggregation of bariatric surgery patients in this project.

In addition, both studies use the predicted values of the size of LDL aggregates at time point t=2h.

Moreover, while both studies look into correlations of LDL aggregation with structure lipids in the LDL lipidome, this thesis also looks into correlations between LDL aggregation and clinical parameters of the patients studied.

As to modelling LDL aggregation, both the original study and this thesis use a slightly modified generalized logistic regression, also known as Richard's curve. This type of curve was chosen because the S-shape it describes is very similar to the shape which data points of LDL aggregation, taken in time, form. Additionally, this curve was originally developed by Richards [34] to model growth, and is widely used in biology. Both works use this curve as a basis for developing a nonlinear mixed-effects model.

Unlike the original study, which uses the Maximum-Likelihood Estimation approach to obtain estimates for the model parameters, this thesis uses the Bayesian modelling approach. One of the reasons for this is because Bayesian modelling allows for incorporating previous knowledge about the parameters through the priors. In addition, it allows for more adequately handling the uncertainty around both parameter estimates and values predicted by the model.

In addition, while the original study assumes a normal distribution for the response variable, which suggests that the variance of all measurements of LDL is the same, in this thesis we choose a Gamma distribution. This allows us to take into account the growing uncertainty of the measurements of LDL aggregates as the values grow larger.

Having covered the basic similarities and differences of the two studies, we now proceed to a discussion of the thesis' methodology.

Chapter 4

Methods

4.1 Data

As already mentioned, this project focuses on studying the LDL aggregation of bariatric surgery patients and its potential correlations with LDL structural lipids, as well as various clinical parameters, collected by the patients. The data was gathered from 33 individuals who underwent bariatric surgery, however, since some of them did not show up for an after-operation follow-up, some of them were excluded, resulting in the final number of patients studied to be 28.

Blood samples and some clinical measurements were taken twice - before the operation, and 6 months after the operation. The blood and plasma samples were then tested, which generated part of the data. The other part were the clinical measurements taken from the patients.

Below we provide details about the available datasets - specifics of the data collection, data cleaning, as well as explanation of the variables they contain.

4.1.1 LDL Aggregation Dataset

As mentioned earlier, the original experiment included 33 individuals who were about to have bariatric surgery. Blood samples were taken immediately before ("BEFORE"-state) and 6 months after the operation ("AFTER"-state). However, in the end some of the subjects did not undergo the surgery or did not attend the six-month follow up. Thus, we only had samples available from their "BEFORE"- state.

All available samples were analyzed using DLS, as described earlier. However, since for some subjects there was only "BEFORE"-data available, these were excluded, as we wanted to focus solely on analysis of the differences of

"BEFORE"- and "AFTER"- states, and keeping these would have been an unnecessary complication. In the end, there were 28 subjects left, for whom before- and after-operation samples were available.

For each subject sample repeated measures of the hydrodynamic radius of LDL aggregated particles were taken throughout a roughly 4-hour time frame. This time frame, as well as the number of measures taken during it, slightly varied for each sample, due to the way in which the DLS instrument works.

Once the DLS data was obtained it was filtered. Only particles which reached a baseline < 1.01 were left. What this means in practice is that the measurements of LDL aggregates with very large radius were excluded, as these are supposed to have considerably more uncertainty, compared to smaller particles.

In addition, a second filter was applied, which excluded particles whose actual correlation function differed significantly from the correlation function fit, measured as sum of squares (SOS). [42]The limit we set was SOS < 100. Again, the goal of this filter is to exclude particles with large fluctuations in their intensity autocorrelation, which is usually a sign of a problem.

In addition, the recommended upper limit of hydrodynamic radius of particles that can be measured accurately by the instrument used, was 1000nm. [43] However, following this recommendation strictly would have left us with very few data points per sample, which would have made modelling LDL aggregation very difficult and, potentially, highly inaccurate. So at first the model we chose was fit with all the data, which resulted in very large standard errors for the predictions made. So several smaller cut-off points for the measurement values were tried out. Eventually, it was decided that only measurements with values of the radius up to 1500nm were to be used. This value was chosen, as it is close to the recommended value of 1000nm, and yet large enough to allow a reasonable amount of data points per sample.

The final LDL aggregation data set contained 368 measurements in total. These measurements were generated from DLS analysis of 56 samples, taken from 28 individuals - 28 samples taken at the "BEFORE" state, and 28 samples taken 6 months after the operation - the "AFTER" state. The final data set contained 5 variables:

• **Pat** - A nominal variable. It contains the id of each patient. This means that two samples in the dataset had the same patient id - a before- and after-sample taken from the same patient.

• **Sample** - A nominal variable. It contains id for each sample taken

• **Time** - A continuous variable. Contains the time points at which measurements of the LDL radius were taken. For each sample the first value of time is 0. And the rest of the time points are relative to it, representing the time which has passed from time point t=0.

• **LDL** - A continuous variable. It contains the values of the hydrodynamic radius of the LDL aggregates.

• **Category** - A categorical variable. It contains two values "B" and "A", which denote the state in which the sample was taken, respectively before and after the operation.

4.1.2 LDL Surface Lipids Dataset

As mentioned earlier, LDL consists of a surface and a core. This dataset contains measurements of the surface lipids present in the LDL particles of the tested subjects, before the onset of LDL aggregation. Again, the data was gathered by first isolating LDL from the plasma and then conducting the analysis of the surface lipids in lab settings. The measure used for the surface lipids is molar percentage, which in this case is defined as the amount of a certain lipid in the surface (measured in moles), divided by the total amount of lipids in the whole surface. The number of surface lipids measured was 49. These constituted 4 main groups:

• **LysoPC** - Lysophosphatidylcholines

• **PCalkyl** - Phosphatidylcholine alkyls

• **PC** - Phosphatidylcholines

• **SM** - Sphingomyelins

As this is beyond the scope of this thesis, the nature of these will not be explained further. As to data cleaning, the data for subjects with only before-operation data available were excluded, as in the first dataset. Two of the surface lipids were excluded, as they had a significant number of 0 values across samples - over 30%. The non-zero values for these lipids were, on average, less than 0.5%. The molar percentages of the surface lipids which were kept were not recalculated to add up to 100% per sample. The rest of the lipids had no value 0, with the exception of two lipids, each of which had four 0 values across samples, but these were kept.

Thus, eventually, the dataset consisted of 48 variables - a sample id and 47 surface lipids which can be found in the surface structure of an LDL particle.

4.1.3 LDL Core Lipids Dataset

This dataset originally contained 30 core lipids - cholesterol esters. As it is beyond the scope of this thesis, we will not discuss the nature of these further, it is sufficient to know that they can be found in the core of an LDL particle. The measure used for the core lipids was molar percentage of the core. To start with, we removed the subjects, for whom we only had before-operation data. Next, as most of those lipids contained more than 50% of zero values, these lipids were removed. The sum of the non-zero values for these lipids for each sample were, on average, less than 0.5% of the total 100% per sample. As in the previous dataset, the molar percentages of the lipids which were kept were not recalculated to add up to 100% for each sample.

Eventually, the dataset contained 8 variables - a sample id and 7 LDL core lipids.

4.1.4 Clinical Dataset

The last dataset contained clinical measurements of the patients taken before and after the operation. As not all of those were relevant to our study, some were removed. In addition, only the data of subjects which had both before- and after-operation data were kept. Eventually, the dataset contained 11 clinical parameters for the patients:

• **Sample**: A nominal variable. Denotes the sample id.

• **Weight**: A continuous variable. Weight of a patient (in kg)

• **BMI**: A continuous variable. Body Mass Index of a patient (kg/m^2)

• **RRSys**: A continuous variable. Average systolic blood pressure of a patient (in mmHg - millimeters of mercury [30])

• **RRDia**: A continuous variable. Average diastolic blood pressure of a patient (in mmHg - millimeters of mercury)

• **Pulse**: A discrete variable. Pulse of a patient (bpm - beats per minute)

• **ogtt0**: A continuous variable. Glucose level at time 0. (in mmol/L)

• **ogtt30**: A continuous variable. Glucose level at 30 mins. (in mmol/L)

• **ogtt60**: A continuous variable. Glucose level at time 60 mins. (in mmol/L)

• **ogtt120**: A continuous variable. Glucose level at time 120 mins.(in mmol/L)

• **HbA1C**: A continuous variable. Average blood glucose levels (in mmol/-mol)

The variables ogtt0, ogtt30, ogtt60, and ogtt120 contain values of the Oral Glucose Tolerance Test (OGTT). This test is used to determine the body's ability to metabolize the intake of sugar. When the test is done, the patient is asked to ingest a glucose drink, and then their blood glucose level is measured at several time intervals. Problems with metabolizing sugar might indicate that the person suffers from diabetes melitus. [3] And since diabetes is one of the risk factors for atherosclerosis, we are interested in these clinical parameters of our test subjects.

As a contrast, HbA1C measures average blood glucose levels, and not the body's short-term ability to metabolize sugar. This clinical parameter is also used as an indicator for diabetes complications [4], which also makes it a quantity of interest for our study.

4.2 Theoretical Background

Before continuing to discussing and comparing the models we built for LDL aggregation, we are going to briefly review some basic theory, related to Bayesian inference and nonlinear mixed-effects models, which is necessary to understand the modelling decisions we make later on. In addition, we discuss leave-one-out cross-validation (LOO-CV), the criterion we use to compare the two models built for LDL aggregation. Additionally, we briefly discuss paired t-test and Wilcoxon rank-signed test, used for testing the differences of the before-after paired values of some variables of interest, as well as Spearman correlation, which we later on use for studying correlations between our variables of interest. In addition we include a brief discussion of the Benjamini-Yekutieli multiple testing adjustment, used to account for testing several hypotheses at the same time. We conclude the chapter by discussing brms - the R package used for fitting our models.

It is worth pointing out that in this subsection we only provide brief

overviews of the concepts and tests used, rather than in-depth discussion. If needed, the reader could find more detailed explanations of the concepts in introductory texts on Bayesian statistics, as well as how and why the tests we use work, in introductory texts on applied statistics.

4.2.1 Basics of Bayesian Statistics

In this subsection we are going to briefly discuss several key concepts of Bayesian statistics. We also include the frequentist view on some of these concepts to serve as a contrast, in order to create some intuition and enhance the understanding of readers who have only been exposed to the frequentist take on statistics.

4.2.1.1 Bayesian Probability

The first important concept we would like to review is the concept of probability. Let us start with the frequentist view on probability.

The frequentist definition of probability is based on considering the outcomes of an experiment which is repeated many times, at least in theory. The probability of an outcome is defined as the limit of the frequency of occurrence of the outcome, considering all possible outcomes, as the number of repetitions of the experiment grows very large. [27]

On the other hand, Bayesians define probability as a quantification of one's personal belief in a certain outcome. This is, how certain a person is, that a certain outcome would take place. Thus, in the context of Bayesian statistics, probability is fundamentally subjective. [27]

4.2.1.2 Bayesian Inference

Another major difference between frequentists and Bayesians is their approach to statistical inference. Statistical inference is the process of drawing conclusions about a population, based on observed data. Drawing conclusions can be done by estimating parameters and predicting new observations. One example of conducting statistical inference is fitting a parametric model to data. Let us now discuss how this is done in both frequentist and Bayesian inference.

Frequentist inference considers model parameters as unknown fixed values, which can be obtained through calculations, based on the observations. The uncertainty around the obtained point estimates for the parameters is expressed through confidence intervals. A confidence interval around a point estimate can only be interpreted as a sequence of repeated similar inferences

about said estimate. [20] One downside to using the frequentist approach to estimating model parameters is that the interpretation of confidence intervals is not very intuitive.

In contrast, while in Bayesian inference parameters are in principle considered fixed, the uncertainty about them is modelled by probability distributions, rather than confidence intervals. [27] This results in Bayesian inference conclusions being presented as distributions for the model parameters and predictions, rather than point estimates, as it is in frequentist inference. One advantage of the Bayesian approach is that one can incorporate some prior knowledge about the value of a parameter, before even considering the data. This prior knowledge can be based on, for example, previous experiments, involving the same type of data, or reflect a personal belief, which makes it, to a large extent, subjective. This prior knowledge about the parameters can then be updated with information obtained from the observed values, which allows for better estimates of the parameters. But, in order to explain how exactly this is done, let us quickly focus on the basis for Bayesian statistics - Bayes' rule.

4.2.1.3 Bayes' Rule

Bayes' Rule, is the basis for Bayesian inference. Bayes' rule is derived by expressing the joint probability distribution of two random variables x, y in two different ways:

$$p(x, y) = p(x|y)p(y) \tag{1}$$

where

p(x,y) - joint probability distribution of x and y

p(x|y) - conditional probability - the probability of x occurring, given that y has occurred

p(x) - the probability of x occurring, independently from y

and also

$$p(x, y) = p(y|x)p(x). \tag{2}$$

Combining these two

$$p(x|y)p(y) = p(y|x)p(x) \tag{3}$$

And rearranging them results in the Bayes' rule:

$$p(x|y) = p(y|x)p(x)/p(y) \tag{4}$$

As mentioned already, Bayes' rule is relevant because it is at the heart of how parameter estimation is done in Bayesian inference. Let us first demonstrate how this is done for a single-parameter model.

Let θ be the parameter of interest, and let $\mathbf{y}$ be the available data. Then, applying Bayes' rule, we get

$$p(\theta|\mathbf{y}) = p(\mathbf{y}|\theta)p(\theta)/p(\mathbf{y}) \tag{5}$$

Each element of the above expression has a specific name and interpretation:

$\mathbf{p(\theta)}$ - is called the prior (probability) distribution of θ. It encapsulates the knowledge available about θ, prior to observing the data $\mathbf{y}$. As mentioned before, this knowledge is considered to be subjective.

$\mathbf{p(y|\theta)}$ - is called the likelihood function. It is the link between the observed data y and the model parameter θ. It encapsulates the information which is contained in the observed data. The likelihood function p($\mathbf{y}|\theta$) is seen as a function of θ, while the observed data $\mathbf{y}$ is considered fixed. [27] The likelihood function is not a probability distribution, as the values the parameter θ can take are not restricted, hence if we integrate over θ (or sum up over, in case of discrete θ), the obtained value would not necessarily integrate (or sum up) to 1, violating one of the requirements for a function to be a probability distribution. [27]

$\mathbf{p(\theta|y)}$ - is called the posterior (probability) distribution. It represents the updated knowledge about the parameter θ after the data $\mathbf{y}$ has been observed. It is one of the most important elements of Bayesian inference.

$\mathbf{p(y)}$ - is known as the evidence or the marginal likelihood, where

$$p(\mathbf{y}) = \int p(\theta)p(y|\theta)d\theta \tag{6}$$

or, in case of θ-discrete:

$$p(\mathbf{y}) = \sum p(\theta)p(\mathbf{y}|\theta). \tag{7}$$

Since p($\mathbf{y}$) does not depend on θ, and the observed data y considered fixed, p($\mathbf{y}$) is considered constant. [20] It serves as a normalizing constant for the posterior distribution, ensuring that it would integrate (sum up) to 1, which

is one of the requirements for a proper probability distribution. [27] p($\mathbf{y}$) is often omitted, thus resulting in the unnormalized posterior distribution:

$$p(\theta|\mathbf{y}) \propto p(\theta)p(\mathbf{y}|\theta) \tag{8}$$

Having explained how the prior and posterior work for a single-parameter model, let us now briefly discuss how they would work in the case of multiple parameters. For the purpose of demonstration, let us assume, without loss of generality, that $\theta = (\theta_1, \theta_2)$, where each of θ_1 and θ_2 can be a vector. In the case of multiple model parameters, the Bayes' rule still holds, but in this case we have a joint posterior density for θ_1, θ_2:

$$p(\theta_1, \theta_2|y) \propto p(\theta_1, \theta_2)p(\mathbf{y}|\theta_1, \theta_2). \tag{9}$$

Let us now assume that we are only interested in the marginal posterior distribution of θ_1, p(θ_1|y). In order to derive this, we simply average over θ_2:

$$p(\theta_1|\mathbf{y}) = \int p(\theta_1, \theta_2|\mathbf{y})d\theta_2 = \int p(\theta_1|\theta_2, \mathbf{y})p(\theta_2|\mathbf{y})d\theta_2[20] \tag{10}$$

The integral above is rarely explicitly evaluated, however it is used to develop a strategy for obtaining the posterior distributions of parameters. These posteriors can be obtained by using marginal and conditional simulation. This is, first θ_2 is drawn from its marginal posterior distribution and then θ_1 is drawn from its conditional distribution, given θ_2. In this way the integral above is evaluated indirectly. [20]

Having discussed the main points on prior and posterior distributions of parameters, it is now worth looking into how parameter estimates are obtained.

4.2.1.4 Parameter estimation

As mentioned earlier, in both frequentist and Bayesian inference, parameters are considered fixed. However, the ways in which parameter estimates are obtained is very different. Again, we are going to first provide the frequentist method of parameter estimation, and then provide the Bayesian approach.

Maximum Likelihood Estimation

In frequentist inference, the most widely-used approach for estimating model parameters is the Maximum Likelihood Estimation (MLE).[27]
In MLE, the estimated value of a parameter is obtained so that it maximizes

the likelihood of observing the given data **y** under the considered model:

$$\hat{\theta}_{MLE} = arg \max_{\hat{\theta}} p(\mathbf{y}|\theta) \tag{11}$$

The result of MLE is a point estimate of the parameter.

Bayesian Approaches to Parameter Estimation

As already mentioned, the emphasis in Bayesian inference is on finding the posterior probability distribution for parameters, hence, naturally, when one is seeking to obtain an estimate for a parameter, this is done through the posterior distribution.

Bayesian parameter estimation approaches, also known as Bayesian estimators, are defined through a minimization problem:

$$\hat{\theta} = arg \min_{\hat{\theta}} \int \int C(\theta - \hat{\theta}) p(x, \theta) dx d\theta \tag{12}$$

where $\hat{\theta}$ minimizes the average cost. [38]
The chosen cost function C(θ) is what determines the approach to parameter estimation. There are several different approaches, but in this paper we discuss in detail only the one used by the R package we used for creating our LDL-aggregation model.

Minimum Mean Square Estimator (The Posterior Mean Approach)

As already mentioned, the parameter estimation approaches in Bayesian inference are defined by the choice of the cost function C(θ). The Minimum Mean Square Estimator is based on the cost function chosen, namely:

$$C(\theta) = \theta^2 \tag{13}$$

Thus, the minimization problem to be solved is then

$$\hat{\theta} = arg \min_{\hat{\theta}} \int \int (\theta - \hat{\theta})^2 p(x, \theta) dx d\theta \tag{14}$$

with a solution

$$\hat{\theta}_{MMSE} = \int \int \theta p(x, \theta) d\theta = E(\theta|x). \tag{15}$$

This is the mean of the posterior distribution, which minimizes the average squared error, hence the name Minimum Mean Square Estimator (MMSE).[38]

The R software package we use for model fitting, brms, uses MMSE to obtain estimates for the model parameters.
Other commonly used estimators are the Maximum A Priori estimator and the Median estimator, which are not discussed further, as they are beyond the scope of this thesis.

Having discussed parameter estimation in Bayesian inference, it is worth discussing how predictions of new values are made in the context of Bayesian inference.

4.2.1.5 Posterior Predictive Distribution

Using the standard notation, let $\mathbf{y}$ be the observed values in an experiment. Let p(θ) be the prior distribution of the parameters.
As already established, the emphasis in Bayesian inference is on finding out the distributions of quantities of interest, rather than simply obtaining point estimates. Thus, if one is interested in predicting a new observation $\tilde{y}$, one can find its distribution, called the posterior predictive distribution, defined as:

$$\begin{aligned} p(\tilde{y}|\mathbf{y}) &= \int p(\tilde{y}, \theta|\mathbf{y}) d\theta \\ &= \int p(\tilde{y}|\theta, \mathbf{y}) p(\theta|\mathbf{y}) d\theta \\ &= \int p(\tilde{y}|\theta) p(\theta|\mathbf{y}) d\theta \end{aligned} \tag{16}$$

As one can see, the second and third row of the equation express the posterior predictive as an average of conditional predictions over the posterior for θ. Note also that the last row follows from the conditional independence of $\mathbf{y}$ and $\tilde{y}$ given θ.[20]

Having discussed the necessary basics of Bayesian statistics, before we delve into finding a suitable model for LDL aggregation, it is important to quickly discuss the general form of regression models. This is important, as it would prove useful later on when we have to take some modelling decisions.

4.2.2 General Form of a Regression Model

In regression modelling the focus falls on the relationship between a response variable y and a set of predictors $\mathbf{X}$, with a function f($\mathbf{X}$) modelling the relationship, i.e

$$y = f(\mathbf{X}) \tag{17}$$

In reality, however, the response variable $\mathbf{y}$ cannot be modelled exactly by f($\mathbf{X}$), as it exhibits a certain level of randomness, i.e $\mathbf{y}$ is a random variable. Hence, the regression model of a relationship between $\mathbf{y}$ and $\mathbf{X}$ is split into two parts:

$$E[\mathbf{y}|\mathbf{X}] = f(\mathbf{X}) \tag{18}$$

$$\mathbf{y} \sim Dist(E[\mathbf{y}|\mathbf{X}]) \tag{19}$$

Where the first expression is the deterministic part, with the function f($\mathbf{X}$) being a deterministic function, modelling exactly the expected value of the response variable $\mathbf{y}$, conditioned on the set of predictors $\mathbf{X}$. The second expression is the stochastic part of the model, which handles the randomness of $\mathbf{y}$. The response variable $\mathbf{y}$ is allowed to vary around its mean, with *Dist* being an arbitrary distribution which models that variation. [27] It is worth pointing out that when we talk about a "distribution" of the response variable $\mathbf{y}$ in this context, we actually mean a distribution family. This is, each observation y_i of $\mathbf{y}$ is such that

$$y_i \sim Dist(E(y_i|X_i)) \tag{20}$$

The type of the distribution for each y_i is *Dist*. From now on in this text the terms "response distribution" and "distribution family of the response" will be used interchangeably.

In summary, choosing a regression model to describe the relationship between a response variable $\mathbf{y}$ and a set of predictors $\mathbf{X}$ means choosing a function f($\mathbf{X}$) and a distribution for $\mathbf{y}$. [27]

Having described the specifics around choosing a regression model, now it is worth focusing on how the distribution of the response variable is chosen. The choice of a response distribution is, to a large extent, based on the nature of the data. In our case we have a continuous response variable, which contains values of measurements of LDL radius that are strictly positive. Hence, we need to look into distributions which incorporate this requirement.

4.2.2.1 Choosing the Distribution Family of the Response

The two distribution families which are most often used in the case of a continuous response variable with strictly positive values, are the log-normal and the gamma distribution families, hence in this thesis we only focus on these two.

Lognormal Distribution

The lognormal distribution is a continuous probability distribution. A continuous random variable X has a lognormal distribution LN(μ_x,σ_x) with μ_x - mean and σ_x - standard deviation, if its logarithm Y = ln(X) Is such that Y $\sim$ N(μ_y, σ_y). [28] The probability density function of the lognormal distribution LN(μ_x, σ_x) is given by

$$f_X(x) = \frac{1}{\sigma_Y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{ln(x)-\mu_Y}{\sigma_Y^2}\right)^2}, 0 < x < \infty \tag{21}$$

where $f_X(x)$ is the probability density function of X, and

$$\sigma_Y = \sqrt{ln\left(\left(\frac{\sigma_X}{\mu_X}\right)^2 + 1\right)} \tag{22}$$

and

$$\mu_Y = ln(\mu_X) - \frac{1}{2}\sigma_y^2. \tag{23}$$

[10] As one can see, the lognormal distribution is only defined for strictly non-negative values, which is what is required in our case.

Gamma Distribution

The gamma distribution is a continuous probability distribution. A continuous random variable X has a gamma distribution with a *shape* a and a *rate* parameter b if:

$$\text{p(x)} = \begin{cases} \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, & x > 0 \\ 0, & otherwise \end{cases}$$

with mean and variance, respectively

$$\text{E[X]} = \frac{a}{b},\ \sigma_X^2 = \frac{a}{b^2}.\ [39]$$

which means that the variance of a variable y_i with a Gamma distribution can be expressed as a function of its mean:

$$Var(y_i) = \frac{1}{b}\mu_i \tag{24}$$

which means that the variance for higher values y_i of the response variable y will be higher, which is what we need in our case, as the uncertainty of the measurements of LDL radius increases as the values of these measurements grow bigger.

In addition, the Gamma distribution is only defined for strictly non-negative values, which is what we need in our case where the response variable contains values of LDL radius, which are all strictly positive.

Having discussed appropriate distribution families for the response distribution, now it is worth discussing a concept, which is closely related to selecting a distribution family - the *link function*.

4.2.2.2 Link Function

The easiest way to introduce the link function is to discuss it in the context of Generalized Linear Models (GLM), hence, we first start with a brief discussion of GLM.

Generalized Linear Models can be seen as an extension to ordinary linear regression, as they allow for flexibility in the choice of response variable distribution. In addition to normal distribution, under GLM a response variable can have a non-normal distribution, such as Gamma, Poisson, etc. [5] In the original formulation of GLM by Nelder and Wedderburn, the distribution of the response variable is a member of the exponential family, however in subsequent works extended GLMs to some non-exponential families. [26]

A GLM has three components:

• The distribution family of the response variable **y**.For simplification of the following explanations, let us consider a single observation of the response variable - y_i.

• A linear predictor $\eta_i = \beta_0 + \beta_1 f_1(x_{i1}) + \beta_2 f_2(x_{i2}) + \cdots + \beta_k f_k(x_{ik})$ where $\beta_0, \beta_1, .., \beta_k$ are unknown parameters and $f_1(.), f_2(.), ..., f_k(.)$ are prespecified functions of the explanatory variables, which can be nonlinear.

• A link function g(.) is a smooth and invertible function which transforms the mean $\mu = E(Y|X)$ to the linear predictor:

$$g(\mu) = \beta_0 + \beta_1 f_1(x_{i1}) + \beta_2 f_2(x_{i2}) + \cdots + \beta_k f_k(x_{ik}). \tag{25}$$

[26]

Since the link function is invertible, one can also write

$$\mu = g^{-1}(\eta) \tag{26}$$

[26] The inverse of the link function is often called *mean function.*

Apart from mapping $E(\mathbf{y}|\mathbf{X})$ to a linear predictor, the link function can be useful in other ways. For example, it could be used to remove restrictions on the range of the expected response μ. [26] For example, let $\mathbf{y}$ be a response variable representing count data. Then it can only take non-negative integer values i.e 0,1,2, ... Let y_i be an observation. Then, the expected count μ_i would also be non-negative (although not necessarily an integer). A log-link function would then map μ_i to the whole real line, thus allowing the parameters that need to be estimated to take any value.

Having discussed the distribution families which we are going to experiment with for our response variable, we now proceed to discuss the next piece of theory which we will later on be applying to model LDL aggregation - nonlinear mixed-effects models.

4.2.3 Nonlinear Mixed-Effects Model

A nonlinear mixed-effects model is a nonlinear model which contains both fixed and random effects. Fixed and random effects can be defined in several ways, some of which are outlined in [19]. The definition we use in this thesis is the following:

• *Fixed effects* are model parameters which are constant across all individuals in a population. [19]

• *Random effects* are model parameters which are allowed to vary across individuals or groups of individuals. [19]

In literature these two terms have been referred to with different names. Fixed effects are also known as *population effects* and *constant effects.* Random effects are also known as *varying effects.* In this thesis, however, we are going to stick to the terms "fixed effects" and "random effects", as these are the most commonly used ones.

The goal of a mixed-effects model is to predict a response $\mathbf{y}$ through a linear combination η of predictors, transformed by the inverse link g^{-1}, assuming a specific distribution *Dist* for the response $\mathbf{y}$. For an observation y_i of the response y, we write

$$y_i \sim Dist(g^{-1}(\eta_i), \theta) \tag{27}$$

The parameter θ describes any additional parameters specific for the distribution family that typically do not vary across observations, e.g standard

deviation sigma of a normal distribution, or the shape parameter alpha of a Gamma distribution. [9]

The linear predictor can be written as

$$\eta = \mathbf{X}\beta + \mathbf{Z}u \tag{28}$$

Where beta are the fixed effects, u are the random effects, and $\mathbf{X}$ and $\mathbf{Z}$ are the fixed- and random-effect design matrices. The response y as well as $\mathbf{X}$ and $\mathbf{Z}$ make up the data, while β, u, and θ are the model parameters that need to be estimated. [9]

As mentioned earlier, the random effects are supposed to vary either across single individuals, or between groups of individuals, defined by grouping factors. For example, one could study the variance of GPA across schools. In this case one can consider school as a grouping factor, and each individual school can be thought of as a group level of this factor.

Let u be the group-level parameters. Then it is assumed that

$$u \sim N(0, \Sigma) \tag{29}$$

[9]

As it is in general the case, the covariances between group-level parameters of different grouping factors are assumed to be zero. Thus, $\mathbf{Z}$ and u can be split up into several matrices Z_k and vectors u_k, where k is the grouping factor index, simplifying the model to

$$u_k \sim N(0, \Sigma_k) \tag{30}$$

[9] In addition, even though not always, group-level parameters associated with different levels, (with index j) of the same grouping factor are usually also assumed to be independent, resulting in

$$u_{kj} \sim N(0, \Sigma_{kj}) \tag{31}$$

It is worth noting that in this case we talk about η as a linear predictor, as this is an easier and more conventional way to explain mixed-effects models. However, it is also possible for η to contain parameters which are nonlinear, which can still be easily handled by software packages used for fitting the mixed-effects model.

Now that we have explained the essence of mixed-effects models, we proceed to the mathematical formulation of LDL aggregation, which is at the focus of this thesis.

4.2.4 Mathematical Model for LDL Aggregation

The mathematical model chosen to describe LDL aggregation is the model used in the original paper. [35] The model is a slightly modified generalized logistic function, also known as Richard's curve. The general form of Richard's curve was first developed by Richard's [34] with the aim of modelling time-dependent growth. It is commonly used in various branches of biology, hence making it a natural choice for modelling of this biological process. Before we proceed to discussing the elements of the LDL aggregation model, it is worth taking a look at how our data "behaves", in order to give the reader some context. The figure below shows LDL aggregation in time for one of the samples in our dataset.

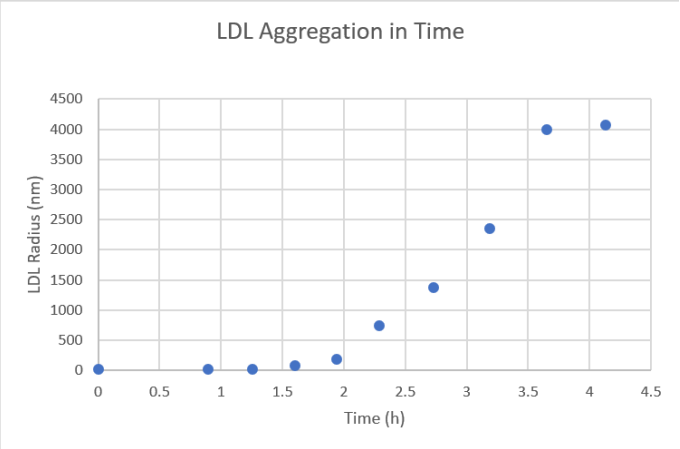

Figure 4.1: LDL aggregation measured in time

As one can see, the data points, representing the size of LDL particles at certain time points, form an S-shape. At the beginning, the aggregation of LDL particles appears to be slower, judging by the size of the aggregates up until around 2h. Then as one can see, the size of the aggregates increases rapidly, forming a steep slope, until it slows down again and reaches a plateau.

This behaviour of the LDL data is modelled as:

$$y_{it} = \beta + \frac{\alpha - \beta}{1 + exp(\frac{\gamma - t}{\delta})} \tag{32}$$

where

• β - the bottom asymptote, obtained when t $\rightarrow -\infty$, which represents the original size of the LDL particles

• α - the top asymptote, obtained when t $\rightarrow +\infty$,which represents the theoretical maximum size of LDL aggregates

• γ - the point of inflection, i.e. the time point where the curve transitions from concave upwards to concave downwards. By definition, this is the point of the curve with the steepest upward slope

• δ - the slope at the inflection point gamma

• **t** - is the explanatory variable which is continuous and represents time. [35]

The model provided above, however, could only be used to estimate the average of the response in different time points, based on the measurements for all patients. However, our goal is to fit a separate model to the data corresponding to each sample, i.e data of a patient in one of the two conditions - before or after the operation. In order to allow for the parameters to vary according to each subset of the data that we fit a model to, we have to use a nonlinear mixed-effects model. The strategy is to calculate fixed effects for the parameters, which would result in a model which gives the average of the response, and then calculate the random effects for some of the parameters, which would result in more flexible models that describe more accurately the behaviour of each subset of data that we fit a model to.

However, the model we introduced in Eq.(32) has four parameters to be estimated, which makes it quite complex. Calculating a fixed and a random effect which correspond to each of these four parameters would make it even more complex. We only have 368 data points, split between 56 subsets, with each subset representing the LDL aggregation data of a certain patient in one of the two conditions - before or after the operation. Our goal is to fit a separate model for each of these 56 subsets, which means fitting a model with four fixed and four random effects based, on average, on less than 7 data points. This means that we need to find ways to simplify the model. As a result, the following modelling decisions were made:

• **Set parameter beta to a constant:** As mentioned earlier, beta is the lower asymptote of the model, which sets a lower limit for the size of LDL particles. After inspecting the lowest value for each of the 56 subsets, we established that none of them goes lower than 15nm, hence, we decided to set beta = 15, and thus simplify the model. And even though for some of the subsets the lowest value was slightly more than 15, setting a common lower limit was a reasonable compromise to make, as the beta-term is just a linear

term and hence does not have such a strong influence on the model.

• **Use only a fixed effect for delta:** In order to simplify the model further, we decided to only use a fixed effect for the slope delta. This decision was made after trying to fit a model with both fixed and random effects for alpha, gamma and delta, which resulted in a model which was too complex and whose parameters could not be estimated.

• **Use both fixed and random effects for alpha and gamma**: We decided to use both fixed and random effects for each of alpha and gamma. The reason for this is because after inspecting the subsets, we established that the highest values for the subsets varied, hence, only selecting a fixed effect for alpha would not have been enough, hence we added a random effect to let the upper asymptote vary across samples. In addition, after a visual inspection of the plots for the separate subsets, we established that the time point where the steep part of each of the plots start also varies, hence, in addition to a fixed effect for gamma, we had to add a random effect to allow the inflection vary across subsets.

It is worth pointing out that here we talk about including random effects for some of the parameters, but we do not make clear what the nature of these random effects is - do they vary across the 56 separate samples (subsets of data points), or per individual, etc. This is in fact an important question which we discuss in detail in the Model Fitting subsection later on. Before continuing to implementing the models, however, we first discuss the model comparison criteria which we use later on to evaluate the developed models.

4.2.5 Model Checking and Comparison Criteria

A thorough model-fitting process would usually involve developing several different models to fit to the same data. Once the models are fit, one needs to perform model checking, in order to ensure that the predictions generated by the models are reasonable.

4.2.5.1 Model Checking

One way to conduct model checking is through posterior predictive checking (PPC). It involves comparing the observed data $\mathbf{y}$ to replicated data generated under the model, usually denoted by $\mathbf{y}^{rep}$. If the model is good, then the data replicated under the model should be very similar to the observed data. [20] Now, let us elaborate on what $\mathbf{y}^{rep}$ is and how it is generated. y^{rep} refers to replicating the experiment which generated $\mathbf{y}$, with the same

values for the explanatory variables **X**, and obtaining as many values as the number of observations in y. The distribution of the replicated data $\mathbf{y}^{rep}$ is defined as:

$$p(y^{rep}|y) = \int p(y^{rep}|\theta)p(y|\theta)d\theta \tag{33}$$

Note that in this case we actually condition on the predictors X as well, but it has been omitted for simplicity. Since this integral is usually difficult to evaluate, the posterior predictive distribution can be computed through simulation:

1. Take N draws $\theta^1, \ldots, \theta^N$ from the posterior distribution p(θ|y).

2. For each draw θ^s, s=1,...,N, an entire vector of simulated values y_s^{rep} is, drawn from the posterior predictive distribution, simulating from the model conditional on the parameters θ^S. [17]

There are several ways to conduct posterior predictive checking. One way is through **visualization**. For example, generating a plot which contains the distribution of y and the distributions of some of the simulated datasets $\mathbf{y}^{rep}$. An example is given in the figure below.

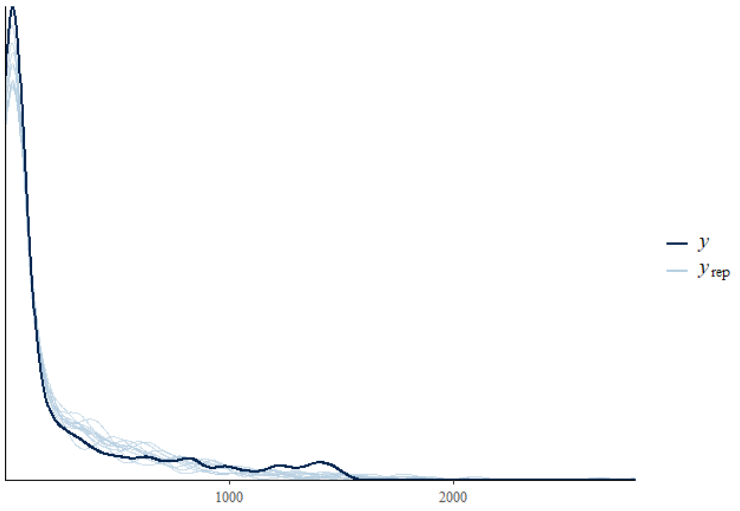

Figure 4.2: Distributions of observations and simulated datasets

Here the thick dark-blue line represents the distribution of the observed data, whereas the thin light-blue lines represent the distributions of the simulated data, calculated for the same values of the explanatory variables as the original data. As one can see, the distributions displayed are reasonably similar, suggesting that the fitted model is reasonable.

In addition, PPC can be done through calculation of **test quantities**. A test quantity T($\mathbf{y}$,θ), also known as *discrepancy measure*, is a scalar summary of the parameters and the data. It is a standard way to compare observed data $\mathbf{y}$ to repeated simulations $\mathbf{y}^{rep}$. The role of test quantities in Bayesian model checking is very similar to the role of test statistics in frequentist testing. When a test quantity is only dependent on the data, it is called *test statistic* and is denoted by T($\mathbf{y}$). In Bayesian context, test statistics can be generalized to allow for their dependence on the model parameters under the posterior distribution. This is a useful way to summarize discrepancies between a given model and the observed data. [20] In practice this is done by calculating the test statistic T($\mathbf{y}^{rep}$) for each simulated set of values $\mathbf{y}^{rep}$, and the distribution of these test statistics is generated. Once this is done, this distribution is then compared to the test statistic T($\mathbf{y}$), calculated from the observed data $\mathbf{y}$. Related to these test statistics, one can compute the posterior predictive p-value, defined as

$$\begin{aligned} p &= Pr(T(y^{rep},\theta) \geq T(y,\theta)|y) \\ &= \int I_{T(y^{rep},\theta)\geq T(y,\theta)} p(y^{rep}|\theta)p(\theta|\mathbf{y})dy^{rep}d\theta, \end{aligned} \tag{34}$$

where I is an indicator function.

What the posterior predictive p-value evaluates is how likely it is for the test statistic value T($\mathbf{y}^{rep}$) for data generated from the model to be more extreme than the test statistic T($\mathbf{y}$). In practice what this means is that an indicator of a good model would be a p-value, which is close to 0.5, while anything too close to 0 or 1 would indicate potential problems with the model. Since this integral is often difficult to evaluate in practice, this can be done through simulation, as described above. First we would draw values for the parameters from the posterior distribution and would then use these to simulate sets of observations $\mathbf{y}^{rep}$. We would then calculate the proportion of cases in which T($\mathbf{y}^{rep}$) > T($\mathbf{y}$). One can also generate a plot of the result, using R:

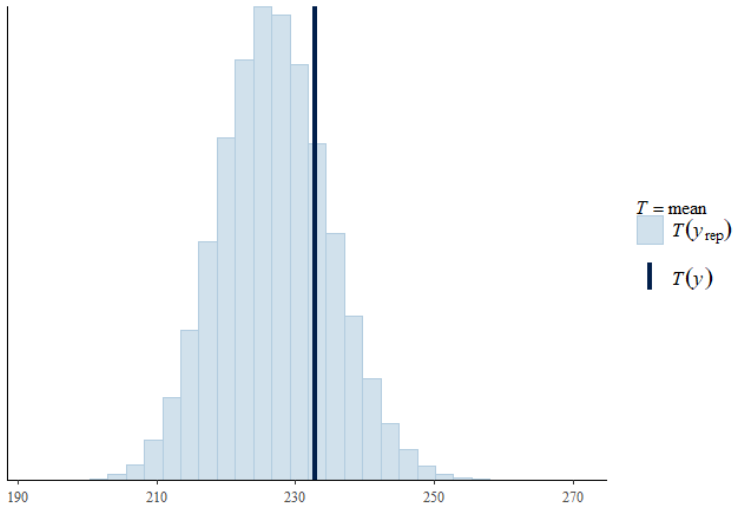

Figure 4.3: Test quantity T(y^{rep}) for the mean, compared to mean of the observations, T(y)

The light-blue histogram is the distribution of T(y^{rep}) values calculated, in this case the mean. The dark line represents the value of T($\mathbf{y}$). If a model is good, then we would expect the mean of this distribution to be as close as possible to the dark line, which is the case in this situation.

Having discussed in detail posterior predictive checking, now we proceed to discussing one of the methods for comparing Bayesian models.

4.2.5.2 Model Comparison

There are several different ways to compare models, based on their predictive accuracy. However in this thesis we will only focus on one of them - Leave-One-Out Cross-Validation (LOO-CV, or just LOO), which we later on use to compare the models we developed.

Leave-One-Out Cross-Validation

Leave-one-out Cross-Validation is a method used for estimating the pointwise out-of-sample prediction accuracy of a fitted Bayesian model. LOO is based on evaluating the log-likelihood at values of the parameters drawn from the posterior distribution. [40]
However, if one wants to conduct LOO in practice, this would mean refitting a model N times (where N is the number of data points in a data set), retaining one of the data points for testing each time, and then evaluating how well the model does predicting it. This makes conducting exact LOO very computationally expensive. Hence alternative ways for evaluating LOO have been developed.

One of the most commonly used ways for approximating LOO is the Pareto-Smooth Importance Sampling (PSIS). PSIS calculates LOO using importance weights, and then fits a Pareto distribution to the upper-tail of the distribution of these importance weights. We are now going to discuss how this procedure works. It is important to note that this is only meant as an overview, so if the reader is interested in the full picture, they should refer to the original paper [40].

The Bayesian leave-one-out estimate of the out-of-sample predictive fit is given by the expected log-predictive density, defined as:

$$elpd_{loo} = \sum_{i=1}^{n} log(p(y_i|y_{-i}))d\theta, \tag{35}$$

where

$$p(y_i|y_{-i}) = \int p(y_i|\theta)p(\theta|y_{-i})d\theta \tag{36}$$

is the LOO predictive density given the data without the point y_i. [40] Provided that the data points are conditionally independent in the data model, one can evaluate the predictive density (36) using values for the parameters θ^s, s=1,...,S, which are drawn from the posterior p($\theta|y$), using *importance ratios*:

$$r_i^s = \frac{1}{p(y_i|\theta^s)} \propto \frac{p(\theta^s|y_{-i})}{p(\theta^s|y)} \tag{37}$$

[40]

These are then used for the Pareto-Smooth Importance Sampling approximation of LOO, which is done in the following way:

1. A generalized Pareto distribution is fit to the 20% largest importance rations r_s, computed above. The importance ratios are computed separately for each held-out point y_i. It is worth noting that the Pareto distribution fit is not sensitive to the cutoff proportion. When the number of draws from the posterior grows very large, the cutoff value should be made smaller. [40]

2. The importance ratios used to fit the generalized Pareto distribution are then stabilized by replacing them with the expected values of the order statistics of the fitted distribution $F^{-1}\left(\frac{z-1/2}{M}\right)$ where z= 1,..., M, and M is the number of importance ratios used to fit the generalized Pareto distribution (M = 0.2S), and F^{-1} is the inverse-CDF of the fitted distribution. These new weights are then labeled as $\tilde{w}_i^s$, where s is the index of the simulation draw and i is the index of the data point. Thus, for each data point y_i there is a vector of S weights, $\tilde{w}_i$. [40]

3. In order to guarantee the finite variance of the LOO estimate, each of the vectors $\tilde{w}_i$ is truncated at $S^{3/4}\bar{w}_i$, where $\bar{w}_i$ is the average of the weights in $\tilde{w}_i$. In this case truncating means that, for a weight w_i^s, we would get w_i^s = min(r_i^s, $S^{3/4}\bar{w}_i$). The truncated weights are then relabeled as w_i^s. [40]

The listed steps are performed for each point. So, for each point y_i we have a corresponding vector w_i, containing S weights. These weights are in general considered to be more stable, compared to the importance ratio which were used to calculate them. [40]
These weights are then used to obtain the LOO estimates. The PSIS estimate of the LOO expected log pointwise predictive density is defined as:

$$\hat{elpd}_{psis-loo} = \sum_{i=1}^{n} log\left(\frac{\sum_{s=1}^{S} w_i^s p(y_i|\theta^s)}{\sum_{s=1}^{S} w_i^s}\right) \tag{38}$$

[40] The reliability of this estimate can be evaluated using shape parameters $\hat{k}$ of the generalized Pareto distributions, which are fit for the raw importance ratios r_i^s, for each data point y_i:

• If for all Pareto distributions fit k < 0.5 , the variance of the raw importance ratios r^s is finite, the Central Limit Theorem (CLT) holds and the

LOO estimate converges fast. [40]

• For k in $[\frac{1}{2}, 1]$, the variance of the ratios r^s is infinite but the mean exists, the generalized CLT for stable distributions holds. In this case the convergence of the estimate is slower. And while the variance of the PSIS-LOO estimate is finite, it is likely to be large. [40]

• For $k > 1$, none of the variance or mean of the raw ratios distribution exist. And, again, while it is finite, the variance of the PSIS estimate is likely to be large. [40]

In practice, observations with $k \leq 0.7$ are still considered reasonable. However, if for a point y_i $k > 0.7$, the estimate of the LOO predictive density given the data without the point y_i (36) cannot be accurately calculated, hence it is recommended that the model is refitted without the point y_i and the LOO predictive density for the problematic point is calculated directly. The points with $k > 0.7$ are called influential points, meaning that they influence the posterior distribution more strongly than the rest of the data points. Having a model with several influential points is quite common, however, if a model has a very large number of influential points, this is usually a sign of a problem with the model.
Comparing models using PSIS-LOO simply means comparing their expected log pointwise predictive density (ELPD). In the case of only two models, as it is the case in this thesis, the model with the bigger ELPD is the one which should be chosen.

Having discussed the basics of Bayesian theory which are used in this thesis, now we move on to discuss the methods which we use to evaluate correlations between our variables of interest, as well as the statistical tests used to investigate change in variables between the before- and after-states.

4.2.6 Methods for Testing Change in Paired Variables

As mentioned earlier, one of the main points of this thesis is to evaluate whether there is a significant change in the variables of interest. This change can be tested in two different ways.

4.2.6.1 Paired t-test

The paired t-test is a standard statistical procedure used to calculate the mean difference between two correlated or dependent sets of observations. [13] In our case we have two sets of measurements of the same variable -

values of LDL aggregation, percentage of a specific surface or core lipid, clinical measurements, taken from a set of individuals before and after bariatric surgery. The fact that these measurements were taken repeatedly - before and after the operation - from the same set of individuals makes the two sets dependent. It is important to point out that both vectors of measurements have the same length and the values contained in the same position of the vectors correspond to the before- and after-measurement of the variable of interest of the same individual, i.e. these are paired observations.

Null Hypothesis: The null hypothesis of the paired t-test is that the mean difference between the aforementioned paired observations is 0.

Assumptions: First, it is important to point out that the paired t-test is used when the two sets contain values of a continuous variable. The use of the paired t-test is based on the assumption that the differences between paired observations are normally distributed. [29] The normality assumption is quite loose, as the test is fairly robust. In cases of extreme non-normality, however, the Wilcoxon signed-rank test is used. [29]

Meaning of p-value: In order to explain how the paired t-test work, let us first briefly explain how it works. Let the number of pairs of observations be n. The difference for each pair of observations is calculated. The mean $\bar{x}$ of these is then calculated. The standard deviation s of the differences is then calculated. These are then used to calculate the test statistic t_s:

$$t_s = \bar{x}/(s/\sqrt{(n)}) \tag{39}$$

[29] The test statistic grows bigger as the mean $\bar{x}$ grows bigger, the standard deviation grows smaller, or as the sample size grows. Then one calculates the probability of obtaining this value of t_s under a true null hypothesis, using the t-distribution - the p-value for the paired t-test. The shape of the t-distribution and thus the probability of obtaining t_s are determined by the degrees of freedom. In this case the number of degrees of freedom is n-1. [29] Applying a paired t-test is, in practice, applying a one-sample t-test to the set of differences between the paired observations. [29]

In case the distribution of the differences between the paired observations is severely non-normal, the Wilcoxon signed-rank test is used.

4.2.6.2 Wilcoxon signed-rank test

Wilcoxon signed-rank test is also used to compare two dependent sets of data, as described in the section about paired t-test above. It is used instead of a

paired t-test when the distribution of the differences of paired observations is heavily non-normal. [29]

Null Hypothesis: The null hypothesis for the Wilcoxon signed-rank test is that the median of the differences of paired observations is 0. It is worth pointing out that it differs from the null hypothesis of the paired t-test, where the mean of the differences was compared to 0.
For the sake of brevity, we omit the explanation of how the test statistic W for the Wilcoxon signed-rank test is calculated, as the steps taken are trivial and readily available in introductory applied statistics literature. [29]

Meaning of p-value: Provided that the data were sampled from a population with a median equal to 0, how likely is it to randomly select N points and obtain a value for the median which is as far or further from 0 as the one obtained from the given set of differences. [24]

4.2.7 Method for Evaluating Correlation

In this thesis we evaluate the correlation between variables of interest using Spearman's rank correlaiton.

4.2.7.1 Spearman's rank correlation

Spearman's rank correlation was chosen as it does not assume that any of the two variables whose correlation is studied are normally distributed or homoscedastic. [29], thus making it less restrictive, compared to the more commonly used Pearson's correlation which assumes normality. In addition, Spearman's correlation does not assume that the relationship between the variables of interest is linear; one can use Spearman's correlation as long as the relationship between the two variables is monotonic [29], thus making it less restrictive and more suitable for the nature of our data.

Null Hypothesis: The null hypothesis for the Spearman's rank correlation is that the correlation coefficient rho is equal to 0, which would mean that there is no association between the variables at hand. [29]

4.2.8 Multiple Comparisons Adjustment

When a large number of statistical tests are conducted at the same time in a study, there arises the possibility of obtaining statistically significant results for some of them, i.e. p-value < 0.05, purely by chance. This problem of

conducting multiple statistical tests at the same time, and as a result rejecting a true null hypothesis due to chance, is known as the *multiple comparisons problem*. Rejecting a true null hypothesis is also known as *Type 1 Error* or a *false positive*. The multiple comparisons problem has been a subject of interest in recent years, due to its importance for all scientific disciplines. [29] Several approaches to controlling for false positives have been developed. Two of the more prominent ones are controlling for Familywise Error Rate (FWER), and finding an acceptable False Discovery Rate (FDR).

Familywise Error Rate is the probability of the set of rejected null hypotheses in an experiment containing at least one falsely rejected null hypothesis. [23] What FWER adjustment methods do is set a level α, denoting the probability of rejecting at least one true null hypothesis, and based on that they impose some adjustments to the set of p-values obtained from the conducted hypothesis tests. Based on the adjustments made for the p-values, usually only a subset of the hypothesis tests which originally yielded statistically significant results are then regarded as significant, corrected for multiple comparisons.
FWER methods are sometimes criticized as being too conservative [23], in the sense that, after the FWER adjustment has been applied, very few results rejecting null hypotheses are considered significant. This could be a problem, as it might result in actually omitting some important scientific result which was rendered insignificant by the multiple comparisons adjustment.

In contrast, **False Discovery Rate** assesses the expected proportion of falsely rejected null hypotheses (false discoveries) in the whole set of rejected null hypotheses. [23] Similarly to FWER methods, a false discovery rate α is chosen, depending on the nature of the problem, which represents the proportion of false discoveries, which are expected to be made. If a follow-up study of a rejected null hypothesis is expensive or can have a huge impact, i.e. implementing a new medical treatment, then, naturally, the false discovery rate is set really low. And, conversely, if a follow-up study is not expensive, one can select a higher α, sometimes as high as 0.2. Most methods developed to control for FWER and FDR assume that the multiple statistical tests conducted are independent of each other. [29] [23] However, in our case, some of the variables about which we form and test hypotheses, have some type of dependency. Under these circumstances, the Benjamini-Yekutieli False Discovery Rate Adjustment is used.

4.2.8.1 Benjamini-Yekutieli False Discovery Rate Adjustment

In our case, some of the variables in our datasets exhibit some type of dependency between each other. For example, we have a dataset of clinical data

which contains the variables weight and Body Mass Index (BMI). The body mass index is defined as BMI = weight/$height^2$, which means that these two variables are correlated. Furthermore, if we conduct hypothesis testing, for example, about the change in weight and BMI before and after bariatric surgery, the test statistics we use would also be dependent. Hence, in this case we require a method which controls for multiple testing and takes into account the dependence of the tests, and as a result the dependence of their respective p-values.

The test we chose, which fulfils these requirements is called the Benjamini-Yekutieli test. It controls the FDR and is valid under a general dependence of the p-values. [23] Similar to other FDR procedures, Benjamini-Yekutieli imposes an adjustment for the p-values, which usually results in only a subset of the rejecting null hypotheses is considered significant. For the sake of brevity, we are not going to explore further how and why the Benjamini-Yekutieli adjustment works. If the reader is interested, they are advised to refer to the original paper. [6]

Another question which arises when multiple comparisons adjustment is to be done, is the question of grouping the hypotheses in meaningful categories, also known as *families*. The reason for that is because multiple comparisons adjustments are usually in some way dependent on the number of hypotheses tested. Testing of very large number of hypotheses together usually results in very strict adjustment which dramatically reduces the number of rejected null hypotheses which are considered significant. One common way of dealing with this problem is to split the set of hypotheses to be tested into groups, based on some underlying similarity. A rule of thumb is to group together the hypotheses which might lead to a single publication. [23] In this way, if one group of researchers conduct an experiment, which generates 100 hypotheses to be tested, and another group conduct a smaller experiment, which only focuses on a testing a subset of these hypotheses, then having some kind of grouping, based on similarity between these hypotheses, allows for a certain level of consistency between the final results. However, this convention is still to some extent arbitrary and does not fully solve the problem.

Now that we have discussed the statistical tests used for testing the hypotheses of interest, it is worth briefly discussing the package used for implementing our LDL aggregation models. While we will try to abstain from going too deep into the technical details, it is worth considering the way in which the programming of the model is done, as it would allow us to put into context the modelling decisions we make in the implementation part.

4.3 Technical Aspects of Model Fitting

In this section we are going to briefly describe the technicalities about fitting our models in R, as well as discuss some convergence diagnostics to ensure that a model has been fit.

4.3.1 The *brms* Package

In order to implement our Bayesian models, we used R statistical language, which is a standard in statistical modelling. And, more specifically, we used the *brms* (Bayesian Regression Modelling using Stan) package. While brms is an R package, it uses the probabilistic programming language Stan on the back-end. Stan is a probabilistic programming language for performing full Bayesian inference. The brms package is an easy-to-use interface for Stan, with excellent documentation and a huge support network in the specialized forum [36], which, from a practical point of view, was another reason to choose the Bayesian approach for this thesis. The brms package uses the No-U-Turn Sampler (NUTS) method, which is an extension of the Hamiltonian Monte Carlo (HMC), a method for drawing random samples from the posterior distribution of developed models. [9] HMC is, in its turn, part of a larger class of algorithms, used for sampling from a distribution - Markov Chain Monte Carlo (MCMC). Compared to other packages which use Metropolitan-Hastings or Gibbs algorithms for sampling, brms has the advantage because the NUTS and HMC algorithms converge much quicker for more complicated models, regardless of whether the prior distributions chosen are conjugate or not [9] which is another reason for choosing it.

Before we continue with fitting the models, we are now going to briefly discuss a piece of code written in R, using the syntax for *brms*, which shows the necessary code to fit a Bayesian nonlinear mixed-effects model.

listings

```
fit_model  <- function(Data){

  fit<- brm(


    bf(
      # (1) Declare the formula formula for the model
      LDL ~ log(15 + (alpha-15) / (1 + exp ((gamma-Time) / delta))),

      # (2) Declare parameters
       alpha ~ 1 + (1|Pat/Cat),
       gamma ~ 1 + (1|Pat/Cat),
       delta ~ 1,
       nl = TRUE),

    # (3) Declare priors
```

```
17
18      prior = c(
19
20      # (3.1) Declare priors for fixed effects
21
22        prior(normal(3000, 2000), class="b",lb=0, nlpar = "alpha"),
23        prior(normal(2.5,2), class="b",lb=0, nlpar = "gamma"),
24        prior(normal(0.2,1), class="b", lb=0, ub=1, nlpar = "delta"),
25
26      # (3.2) Declare priors for other distribution parameters
27        prior(gamma(0.01, 0.01), class="shape"),
28
29      # (3.3) Declare priors for SD of random effects
30
31        prior(normal(0,2), class="sd", group="Pat:Cat", coef="Intercept",
              nlpar="gamma"),
32        prior(normal(0,1000), class="sd",group="Pat:Cat", coef="Intercept",
              nlpar="alpha")),
33
34      data = Data,
35
36      #(4) Declare family distirbution for the response variable and link-
            funcition
37      family = Gamma(link="log"),
38      chains = 4,
39      iter=50000,
40      warmup = 1000,
41      cores = getOption("mc.cores",1L),
42      thin = 1,
43      control = list(adapt_delta = 0.99999, stepsize=0.00001, max_treedepth
            =15)
44    )
45
46    return(fit)
47  }
```

Let us now go through the main steps marked with comments in the code above:

1. **Step 1:** This is where the model formula is declared.

$$\textbf{LDL} \sim \textbf{log(15 + (alpha-15) / (1 + exp ((gamma-Time))))}$$

On the left-hand side we have the response variable, in our case LDL, and on the right hand side we have the model formula. Note that in this case we have the original formula for our LDL aggregation model, which is inside a log() function. This is because in this example the chosen link function is the log-link and we want

$$\begin{aligned}\textbf{E(LDL|X)}&\textbf{= 15 + (alpha-15) / (1 + exp ((gamma-Time) / delta))}\\ &\textbf{= exp(log(15 + (alpha-15) / (1 + exp ((gamma-Time))},\end{aligned}$$

where exp() is the inverse-link function.

2. Step 2: In this step we declare the parameters as a combination of fixed and random effects. In this case, for example, we have

$$\textbf{alpha} \sim \textbf{1} + \textbf{(1|Pat/Cat)}$$

Which means that for the parameter alpha we have a fixed effect and we also have 2 random effects - one which varies between patients (1|Pat), and one effect which is within the group factor Pat (patient), which varies between the two categories of Pat (B - before and A-after).

3. Step 3: This is where the priors are set.

3.1. Priors for the parameters
This is where the priors for the parameters are set. Here one can also set upper and lower bounds which make sense for the parameters in the specific context.

3.2. Priors for response distribution parameters
These are usually default, so the user does not need to worry about them.

3.3. Priors for the SD of random effects
As already mentioned, random effects are assumed to be normally distributed with mean 0, so the prior is used to evaluate the standard deviation.

4. Step 4: This is where the family distribution and the link-function are declared.

Now that we have discussed how a model is declared, we are going to briefly talk about convergence diagnostics for the sampling algorithm, used to approximate the posterior distrobution.

4.3.2 Convergence Diagnostics

As mentioned earlier, the *brms* package uses the NUTS sampling algorithm, which is used to draw samples from the posterior distribution. Briefly, the way this sampler works is, it starts off from some initial value and, based on a set of rules, makes further draws (iterations), with the idea that at some point the draws it makes are actually from the posterior distribution. When the algorithm gets to the point of drawing samples from the posterior distribution, we talk about *convergence* to the posterior distribution. For the sake of brevity, we are not going to elaborate further on how this works in

theory, but one can easily look this up in the literature. Various diagnostics have been devised to evaluate convergence.

To start with, a sampling algorithm is usually run on multiple different *chains*. This is, it is ran in parallel for several different initial values. The algorithm is considered to converge if eventually the values drawn from each chain are quite similar.
The "similarity" is measured by the value $\hat{R}$, known as *potential scale reduction factor*. For each parameter of interest, the variance for each chain is estimated (after discarding the initial M draws from each chain, as these are considered to be far away from values in the posterior). Then the average of these within-chain variances is calculated. Additionally, the variance of all chains mixed together is calculated. The calculated $\hat{R}$ is the square root of the mixed variance, divided by the average of the within-variances. When convergence has been reached, all chains will have mixed, and thus the distribution of the draws between and within chains will be identical. Which means that an $\hat{R}$ would be close to 1 (in practice values less than 1.1 are considered to be indicating convergence). [21]

In addition, when evaluating the convergence to the posterior, another factor to look at is the *effective sample size*(ESS). Ideally, samples drawn by a sampling algorithm would be independent of each other. However, this is usually not the case. The existing autocorrelation within chains increases the uncertainty of the estimates. This increase of uncertainty can be measured by using ESS. When independent samples are considered, the uncertainty of estimates is bound to the number of samples N. When considering dependent samples, N is replaced by the effective sample size N_{eff}, which is defined as the number of independent samples, which have the same estimation power as the autocorrelated samples. [2] This means that the smaller N_{eff}, the greater the uncertainty about the obtained parameter estimates is. As a rule of thumb, an N_{eff} smaller than 10% of the number of total iterations run is alarming and indicates that the chains of the sampling algorithm should be run longer, in order to converge.

We now proceed to fitting the developed models, where we will be discussing the configurations chosen in the context of the 4-step plan discussed above.

4.4 Model Fitting

In this section we are going to discuss two different nonlinear mixed-effects models which were fit to model LDL aggregation, do model checking and model comparison on these models.

As noted earlier on, the parameter β was set to a constant, $\beta = 15$, in order to simplify the model. In addition, it was decided that the parameter δ would be a fixed effect. Hence, the possible variations of the model would, in essence, be centered around experimenting with random effects for α and γ. In addition, as mentioned earlier, the idea of mixed-effects models is to allow more flexibility of the fitted models, with the values of the model parameters varying across individuals and groups. In our case we have 56 subsets of data, each of which containing LDL aggregation data for a specific patient in one of the two conditions - before- or after-operation. For brevity, in the next section we will refer to these subsets as *samples*

4.4.1 Model 1

4.4.1.1 Model Declaration

As stated above, we will describe this model based on the four steps suggested above.

1. Model formula

The model formula for this model is the one stated in (32).

2. Fixed and random effects related to parameters

Using the notation of brms, the model parameters are defined as:

$$\begin{gathered}\alpha \sim 1 + Cat + 1|Pat \\ \gamma \sim 1 + Cat + 1|Pat \\ \delta \sim 1.\end{gathered}$$

In the case of α and γ we have two fixed effects and one group-level random effect.
The first fixed effect is shared by all individuals. The second effect is related to the variable *Category*, denoted by Cat. This variable has two levels - B-before and A-after, corresponding to before- and after-operation states of the patients. What this fixed effect does is it serves as a *difference* between the before- and after-condition of each patient, and is constant across patients. An estimate of this fixed parameter is obtained for only one of the levels of the variable - in our case CatB. In our case estimates α_{CAT_B} and γ_{CAT_B} are obtained, and these are then added to, respectively, the α- and γ-values only of *samples*, representing a patient in a before-operation state. Note that the value of α_{CAT_B} and γ_{CAT_B} is the same for all individuals in before-operation

state. Note that the idea of this fixed effect is actually consistent with the idea of a non-zero, i.e. significant average difference between before- and after-operation values of LDL.

The group-level random effect $1|Pat$ varies across patients. This means that the models for the samples, related to the same patient, would have the same value value for the random effect $1|Pat$.

3. Priors

3.1. Priors for fixed effects

As already mentioned, all three model parameters have a fixed effect associated with them. The priors we chose for these fixed effects are, respectively:

$$\alpha_{fixed} \sim N(3000, 2000)$$

The reason why this distribution was chosen is because we do not have much information about this parameter, and a normal distribution is typically used in this case. Such distribution, which does not provide a lot of insight into a parameter, but still incorporates some level of knowledge, is usually referred to as *weakly informative*. Based on previous experiments and advice from a domain expert, we know that usually this type of LDL aggregation data grows and settles at a plateau which is somewhere around 3000, hence the mean we chose is 3000. As one can see, however, the standard deviation is set very high, to capture the high level of uncertainty around α.

In addition, we have the fixed effect for α, related to Category, α_{CatB}.

$$\alpha_{CatB} \sim N(0, 1000)$$

Since we have no information about the distribution of this effect, we choose a normal distribution with mean 0 and a large standard deviation, 1000, to capture the large level of uncertainty around this α_{CatB}.

For the fixed effect for gamma, γ_{fixed}, we have

$$\gamma_{fixed} \sim N(2.5, 2)$$

Similarly to α, we chose a normal distribution for γ as well. The mean was chosen to be 2.5, as based on knowledge from previous experiments and domain expert's opinion, we know that the inflection point for LDL aggregation data of this type is usually around 2.5h. However, in order to

reflect the high level of uncertainty, we chose a large value for the standard deviation, in this case $\sigma = 2$.

We also set a distribution for the fixed effect γ_{CatB},

$$\gamma_{CatB} \sim N(0, 2)$$

Again, due to lack of information about this fixed effect, we choose a normal distribution, with mean 0 and a standard deviation 2, which is fairly large, considering the range of values which the variable time takes.

Similarly for δ,

$$\delta_{fixed} \sim N(0.2, 1)$$

the normal distribution was chosen. The mean was chosen to be 0.2, based on past experience, where the slope at the inflection point was around 0.2. Again, the standard deviation was set quite high, in order to reflect the high level of uncertainty, $\sigma = 1$.

3.2 Priors for response distribution parameters

The prior for the shape of the Gamma distribution, chosen as a family distribution, is assigned, by default by the brms package, to be

$$\text{shape} \sim \text{Gamma}(0.01,\ 0.01)$$

Where the first argument is μ - mean, and the second one is the shape. Again, this prior is very vague, but it is assigned purely to stabilize the model to be fitted.

3.3 Priors for the standard deviation of random effects

In this case we have two random effects, one for α and one for γ. As already mentioned, the distributions declared are for the standard deviations of the random effects. These are:

$$sd_{\alpha_{1|Pat}} \sim N(0, 1000)$$

Again, this prior is deliberately chosen to be weakly informative, as we do not have information about this random effect. Since we don't have any information, we chose to set the mean to 0, and chose a large value for the standard deviation to reflect the high level of uncertainty . Similarly, we chose

$$sd_{\gamma_{1|Pat}} \sim N(0, 2)$$

which is again a prior which does not make any strong assumptions, as we do not have much information about this random effect. The mean is again 0, and the standard deviation is chosen to be 2. Note that while this is significantly smaller than the standard deviation for the other random effect, in the context of the units in which gamma is measured - hours, it is still quite vague and reflecting the large level of uncertainty.
It is worth pointing out that, since all the priors we chose are weakly informative, the observed data would have a stronger influence on the model. This is, by setting weakly informative priors we "let the data speak for itself".

4. Response distribution family and link-function

Family: Gamma()
Link function: Log-link

The family for the response distribution which is chosen is the Gamma family. The reason for this is because as mentioned earlier, in Gamma distribution the variance increases with the mean. This is relevant to our case, as the uncertainty of the measurements of LDL aggregation grows as the values of the measurements increase. Hence, this property of the Gamma distribution adequately reflects the nature of our data. In addition, we also experimented with the lognormal family, but the resulting model was quite a poor fit to the data, hence this option was not explored further.

As to the link-function, the log-link was chosen. This link function was chosen through experimenting. As the predictor resulting from our model does not have a nice, intuitive interpretation, as it is nonlinear, hence an experiment resulting in a good model is our strongest argument for using the log-link in this case. In addition, we experimented with the identity link function, but this resulted in a model, which did not converge, meaning that we could not obtain reliable estimates for our model parameters. We also experimented with the inverse-link function, which yielded results which were very similar to the case when log-link was used, hence we decided to stick to the log-link, purely for convenience.

4.4.1.2 Model Evaluation

1. Convergence Diagnostics

As one can see in Appendix A.1.1 all parameters have $\hat{R} = 1.00$, which suggests that the sampling algorithm has converged to the posterior distribution. In addition, all effective sample sizes are reasonably large, which

means that it is likely that the samples drawn by the algorithm are reasonably close to the posterior distribution.

2. Model Checking

In order to evaluate the posterior predictive performance of Model 1, first we are going to compare the distribution of observed values to several simulations of predicted values, calculated for the same respective values of the explanatory variables, as discussed in Subsection 4.2.5.1. The corresponding plot is given below.

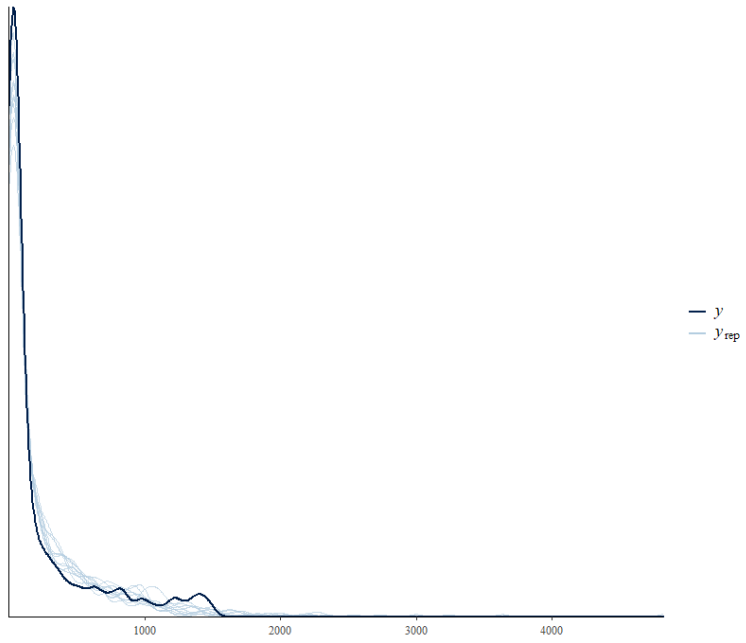

Figure 4.4: Distribution of observed values, compared with distributions of simulated values drawn from the posterior predictive distribution, using Model 1

As one can see, the dark-blue line, representing the distribution of the observed values, has quite similar shape to the light-blue lines, representing the distributions of the simulated values drawn from the posterior predictive distribution. This suggests that the model fitted is doing reasonably well is posterior prediction.
In addition, we calculate test quantities, discussed in Subsection 4.2.5.1. In our case we decided to calculate both the test quantity T1 - mean and T2 - skewness, as these the combination of these two is an adequate way to de-

scribe a distribution. The skewness function used to obtain this quantity is known as Pearson's second coefficient [41], defined as:

$$Sk_2 = 3\frac{(\mu - \nu)}{\sigma} \tag{40}$$

where μ - mean, ν - median, and σ - standard deviation. The plot illustrating these test quantities is given below:

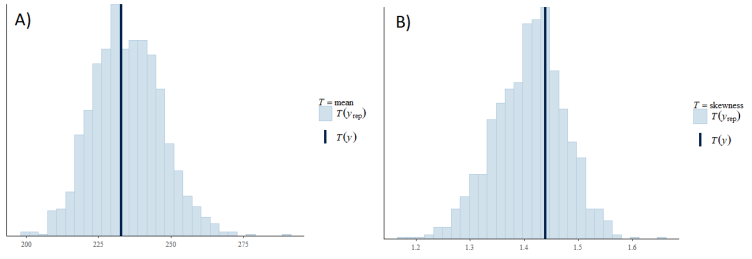

Figure 4.5: Test quantities: A) T_1 - mean and B) T_2 - skewness for Model1

As one can see, in both cases the values for $T_1(y)$ and $T_2(y)$ are quite close to the mean of the distributions, generated from calculating these test quantities for the simulated values y^{rep}, which suggests that the model performance in terms of posterior predictive power is reasonable.
We now continue to the second model we fitted.

4.4.2 Model 2

4.4.2.1 Model Declaration

1. Model formula

The model formula for this model is the one stated in (32).

2. Fixed and random effects related to parameters

Using the notation of brms, the model parameters are defined as:

$$\begin{gathered}\alpha \sim 1 + 1|Pat/Cat \\ \gamma \sim 1 + 1|Pat/Cat \\ \delta \sim 1.\end{gathered}$$

This means that there is a fixed effect and two random effects for both α and γ, and a fixed effect for δ. But before focusing on the priors chosen, let us

first explain the two random effects.
These are explained in the expression

$$1|Pat/Cat,$$

which can be expressed as

$$1|Pat + 1|Pat : Cat.$$

The random effect $1|Pat$ describes the variation across patients (Pat). In addition, the random effect $1|Pat : Cat$ describes variation across the group-factor of patient (Pat) and its group-levels, determined by category (Cat). We now proceed to discussing the priors.

3. Priors

3.1. Priors for fixed effects

The priors for the fixed effects $\alpha_{fixed}, \gamma_{fixed}$, and δ are the same as the ones chosen for *Model 1.*

3.2. Priors for response distribution parameters

The prior for the shape of the Gamma distribution, which was again chosen as a family distribution, is, again, assigned by default by the brms package.

3.3 Priors for the standard deviation of random effects

The distributions chosen for the standard deviations $sd_{\alpha_{1|Pat}}$ and $sd_{\gamma_{1|Pat}}$ are the same as for *Model 1.* In addition, however, we also have two more prior distributions for the standard deviations of the random effects of α and γ, respectively:

$$sd_{\alpha_{1|Pat:Cat}} \sim N(0, 1000)$$

and

$$sd_{\gamma_{1|Pat:Cat}} \sim N(0, 2)$$

As one can see, once again we opted for the normal distribution for the priors of these standard deviations. Again, the reason for this is because we do not have any concrete prior knowledge about their values. Hence we set up priors which only reflect our vague idea of the scale of these values, captured mainly by the standard deviations of these normal distributions. Yet, even though

very weakly informative, these priors ensure some computational stability, by imposing some reasonable limits of the search space for the values of interest.

4. Response distribution family and link-function

The response distribution family and the link-function are the same as chosen in *Model 1*.

4.4.2.2 Model Evaluation

Convergence Diagnostics

As one can see in Appendix A.1.2 all parameters have $\hat{R} = 1.00$, which suggests that the sampling algorithm has converged to the posterior distribution. In addition, all effective sample sizes are reasonably large, hence, it is reasonable to assume that the draws of the algorithm are reasonably close to values drawn from the posterior.

2. Model Checking

In order to evaluate the posterior predictive performance of Model 2, first we are going to compare the distribution of observed values to simulations of predicted values, calculated for the same respective values of the explanatory variables. The corresponding plot is given below.

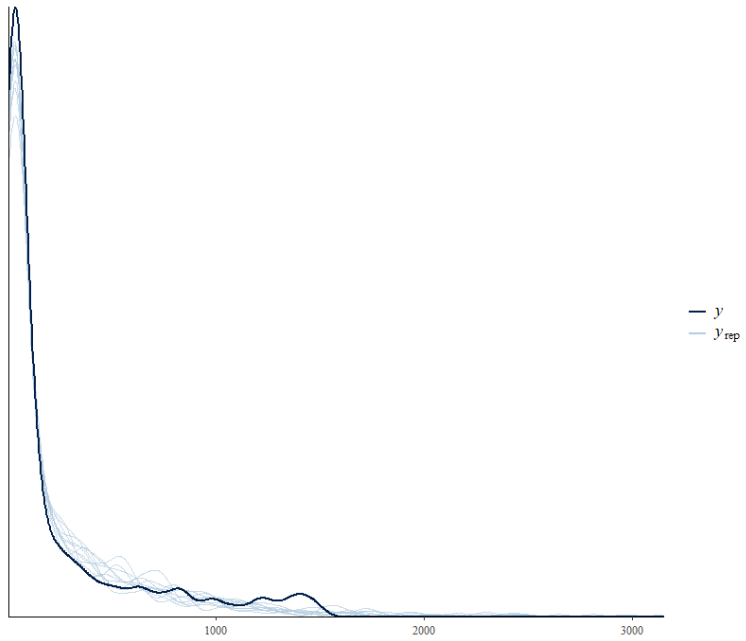

Figure 4.6: Distribution of observed values, compared with distributions of simulated values drawn from the posterior predictive distribution, using Model 2

As one can see, the dark-blue line, representing the distribution of the observed values has a very similar shape to the light-blue lines, representing the distributions of the simulated values drawn from the posterior predictive distribution (for the same respective values of the explanatory variables). This suggests that the model fitted is doing reasonably well is posterior prediction.

We now calculate the test quantities we calculated for Model 2, T1 - mean and T2 - skewness. The results are provided in the plot below.

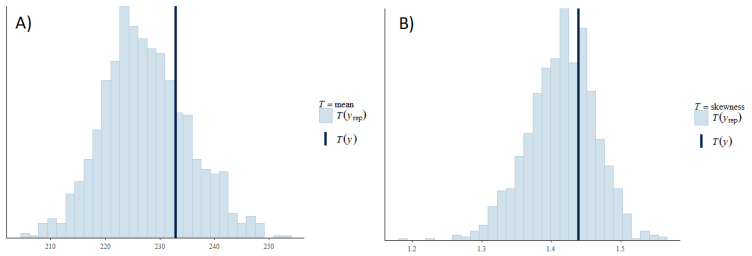

Figure 4.7: Test quantities: A) T_1 - mean and B) T_2 - skewness for Model2

As one can see, $T_1(y)$ is not as close to the mean of the distribution of $T_1(y^{rep})$-values for Model 2, as it is for Model 1. Still it is reasonably close. In plot B) $T_2(y)$ is very close to the mean of the distribution of $T_2(y^{rep})$-values. Based on both plots, it seems reasonable to assume that Model 2 does fairly well in posterior prediction.

4.4.3 Model Comparison

Now that we have fit two models which seem to perform well in the posterior predictive checks, we are going to compare them. We are going to use the Leave-One-Out Cross-Validation, described in Subsection 4.2.5.2, which evaluates the pointwise out-of-sample prediction performance of the models.

In order to do this comparison, we run a function in *brms* package which calculates the approximation of the LOO expected log predictive density (elpd), using the Pareto-Smoothing Importance Sampling (PSIS), as described in Subsection 4.2.5.2. However, as explained earlier, there are cases in which the log pointwise predictive density (lppd) cannot be adequately approximated for some data points. If the number of these influential points is reasonably small, this usually means they are just outliers, which the model fitted is "surprised" to encounter. The diagnostic used to determine whether a point is influential is the pareto-k ($\hat{k}$) parameter, as discussed earlier. Points with pareto-k > 0.7 are considered problematic. We generated pareto-k plots for both of our models, provided below, provided in Appendix A.2. As one can see from Figure A.1, Model 1 has 6 influential observations. From Figure A.2 one can see that Model 2 has 7 influential observations. Thus, we refit each of the models, retaining one of the influential points at a time. Once these are refitted, we are able to calculate the elpd for each of the models and compare them. The output of the comparison is given below.

```
1   elpd_diff se_diff
2 model2    0.0       0.0
3 model1 -122.0      21.1
```

This output means that Model 2 has a higher expected predictive accuracy. Thus, we choose to use Model 2 for the prediction of the LDL values at 2h-time point.

4.4.4 Performance of Model 2 on a test set.

Now that we established that Model 2 is the better of the two models, in order to ensure that our predictions of the 2h-values would be reasonable, we decided to test it on an existing test set. Since we do not have any LDL values for time point exactly 2h, we decided to test our model on the closest available ones. For all samples we have a measurement of LDL, taken around time point t=1.94h, which is very close to 2h. Thus we retained 10 LDL values measured at t = 1.94h, and then refit Model 2 without them. Then we tested its predictive performance on these values. The mean absolute error of the predictions was **MAE=0.2386**. While it is desirable for this value to be smaller, it can still be considered reasonable, considering that the data set used to fit the model is fairly small. We then proceeded to obtaining the predicted values of LDL at the 2h-time point.

Having obtained these, we now proceed to discussing the results from the statistical tests we conducted.

Chapter 5

Results

In this section we state the results obtained in this thesis project. The results, presented below, are given as answers to the main research questions asked in the thesis, which were stated earlier on in the Introduction section. Throughout the whole result section, when we talk about LDL values, we mean the radius of LDL aggregates at time point t = 2h, which we use as a single quantitative measure of LDL aggregation.

1. Is there a significant difference in LDL aggregation of bariatric patients before and after the operation?

The LDL aggregation values are obtained for each one of the 56 samples, by using the fitted LDL aggregation model to predict them. Thus, we have available 28 before- and 28 after-operation LDL 2h-time point values, which correspond to the 28 individuals we have data for. We then form differences between the pairs of values of each patient. The difference is formed in the direction before minus after values.
In order to establish whether there is a statistically significant difference between before- and after-operation aggregation of LDL, we use the paired t-test.

Null Hypothesis: The mean difference between the before-after pairs of LDL aggregation values is 0.

Significance Level: The significance level we choose for the paired t-test is the standard in the scientific community, p-value = 0.05.

Checking Assumptions: We assume that the differences of before-after pairs are independent of each other, as we assume that the measurements

taken for each person are independent. As to the second assumption of normal distribution of the differences, we generate the following histogram:

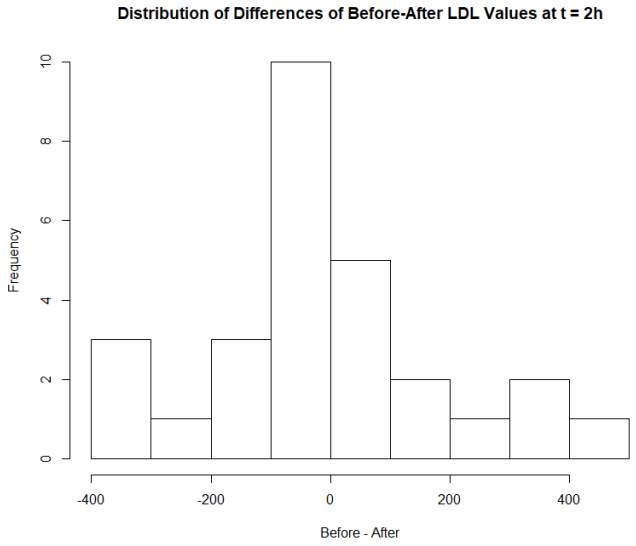

Figure 5.1: Histogram of differences between before and after LDL values at t=2h

As one can see, the distribution of before-after differences appears to be roughly normal, which is sufficient, as the paired t-test is fairly robust. Hence, we proceed to applying the paired t-test to formally test the mean difference of before-after LDL aggregation. The result is given below.

Result:
The R output of the paired t-test is the following:

```
Paired t-test
data:  before_ldl and after_ldl
t = -0.3815, df = 27, p-value = 0.7058
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -90.14508  61.87912
sample estimates:
mean of the differences
              -14.13298
```

As one can see, the sample estimate for the mean of differences
μ = **-14.13298**, but the **p-value = 0.7058**, hence, this is not a statistically

significant result on the 0.05-level. In addition, consistent with the conclusion indicated by the p-value, the 95%-confidence interval contains 0.
Thus we fail to reject the null hypothesis that the mean difference is 0.

2. Are there any significant changes in the structure of LDL particles of the bariatric patients before and after the operation?

Before we continue with answering this question, let us first discuss what exactly is asked. In the context of this question "changes in the structure of LDL" means changes in the amount of the various lipids of LDL before and after the operation. Thus, in order to investigate potential changes in the lipids, we are going to form differences between their before- and after-values.

As already mentioned, the structure of LDL particles can be divided into a surface and a core. Thus, in order to answer the question, we are going to separately investigate changes in surface lipids and changes in core lipids. As we are going to test multiple hypotheses for each of these cases, we are going to use the Benjamini-Yekutiel multiple comparison adjustment and consider these as two separate families, respectively - surface-lipids and core-lipids family of hypotheses.

2.1. Change in Surface Lipids

In order to investigate changes in surface lipids, we form the difference between the before- and after-operation values for each lipid. We then generate histograms of the differences between the before and after pairs for each lipid. For the lipids whose differences appear to follow a roughly normal distribution, we use paired t-test for formally testing the mean difference between the before and after values. For the lipids whose differences appear to follow a clearly non-normal distribution, we use the Wilcoxon signed-rank test for formally testing whether the median of the differences is 0.

Null Hypotheses: Since the differences of lipids have to be tested with two different tests, it is worth pointing out that the null hypotheses that are being tested are different.

• **Null Hypothesis for Paired T-test:** The mean difference between the before-after pairs of surface-lipids values is 0.

• **Null Hypothesis for Wilcoxon Rank-Order Test:** The median of the differences between the before-after pairs of surface-lipids values is 0.

Significance Level: The significance level we choose for both the paired and the Wilcoxon Rank-Order t-test is p-value = 0.05.

False Discovery Rate Level: In order to adjust the results for multiple comparisons, we use the Benjamini-Yekutieli adjustment. The FDR level chosen is $\alpha = 0.1$. This value is standard for FDR multiple comparison adjustments.

Result: For the sake of brevity, we only provide the results which were deemed significant after the Benjamini-Yekutieli multiple comparisons adjustment. The results are split into two tables - one containing the statistically significant results obtained from paired t-tests and the other one containing the results obtained from Wilcoxon rank-order tests. It is important to point out the distinction, as the two tests have different null hypotheses. The two tables are given below.

Paired T-test Results				
Surface Lipid	**Mean Diff.**	**95% CI**	**p-value**	Avg.% Change
LysoPC18:00	0.1182143	(0.0573, 0.1791)	0.000463	0.067
LysoPC18:02	0.06357143	(0.0194,0.1078)	0.00649	0.2
LysoPCtotal	0.3453571	(0.1357, 0.5550)	0.00222	0.11
PC34:01alkyl	-0.06285714	(-0.0893,-0.0364)	< 0.0001	0.23
Pcalkyl total	-0.3032143	(-0.4880,-0.1185)	0.0023	0.1
PC32:00	-0.1018	(-0.1355,-0.0680)	< 0.00001	0.22
PC34:02	1.605	(0.7124, 2.4976)	0.001	0.09
PC36:02	1.2593	(0.7002, 1.8183)	< 0.0001	0.11
PC38:03	0.4025	(0.1769,0.6281)	0.001077	0.19
PC40:07	-0.0661	(-0.0895,-0.0426)	0.00001	0.42
SPHM16:00	-1.0036	(-1.3852,-0.6219)	<0.00001	0.15
SPHM21:00	0.0936	(0.0244,0.1628)	0.009893	0.34
SPHM23:00	0.2382143	(0.1508, 0.3256)	< 0.00001	0.24
SPHM24:01	-1.4704	(-1.9196,-1.0211)	<0.000001	0.28
SPHMtotal	-2.5675	(-3.8045,-1.3305)	0.000223	0.13

Table 5.1: Surface lipids which have a mean difference of before- and after-operation values, which is different from 0 and statistically significant, measured by paired t-test

The Average percentage change (Avg. % Change), provided in the table,

is calculated as the absolute value of the mean difference, divided by the average value the specific lipid for the before-operation state(given in molar percentage).

Wilcoxon Signed-Rank T-test Results			
Surface Lipid	**Median Diff.**	**95% CI**	**p-value**
PC36:01alkyl	-0.0350	(-0.0501,-0.0150)	0.005186
Pctotal	2.34	(1.095, 3.740)	0.000425
SPHM24:02	-0.57	(-0.92, -0.25)	0.000561

Table 5.2: Surface lipids which have a median of the differences between before- and after-operation values which is different from 0 and statistically significant, obtained by Wilcoxon signed-rank test.

The number of rejected null hypotheses before adjusting for multiple comparisons was in total 27, but after the adjustment decreased to 18, which is a significant change.

2.2. Change in Core Lipids

In order to investigate the change in core lipids, we inspected the differences of the pairs of before- and after-operation values for each lipid. The null hypothesis, significance level, and FDR level are the same as in Question 2.1. The differences for all core lipids were normally distributed, hence we used the paired t-test in all cases, in order to formally test the difference. Initially the core lipids ChoE18:01 and ChoE18:03 were found to have a mean difference which was different from 0 and statistically significant. However, after applying the Benjamini-Yekutieli multiple comparisons adjustment with FDR level α=0.1, none of the results were deemed significant.

3. Are there any significant correlations between LDL aggregation and LDL structure lipids before and after the operation?

As in the previous question, we needed to investigate correlations between LDL and surface and core lipids, respectively, by considering them separately. In all cases considered, the correlation test which was used was the Spearman's correlation test, with p-value < 0.05 chosen as a significance level. The Benjamini-Yekutieli multiple comparisons adjustment was used, with FDR level α=0.1 in all cases.

3.1. Correlations between LDL Aggregation and Surface Lipids

In order to answer this question, we considered separately the pair of before-operation values for LDL and the surface lipids, and the pair of after-operation values, respectively. The reason why we could not look into a correlation of all LDL values (both before- and after-) and all values of each lipids, was because the before- and after-values for LDL, as well as for all lipids, are correlated. They are correlated because each before-after pair corresponds to the same patient, thus violating the independence-of-observations assumption, required when calculating correlations.
We split the hypotheses tested to Before-correlations and After-correlations families, respectively. We applied the Benjamini-Yekutieli adjustment to the p-values obtained from the Spearman correlation tests.

a) Correlations between Before-values for LDL and Before-Values for Surface Lipids

The initial test results suggested that the lipids LysoPC16:00, LysoPC20:04, as well as the total sum of the molar percentages of all LysoPC-lipids (LysoPC-Total) for each patient, had a statistically significant correlation with the before-values of LDL. However, after applying the multiple comparisons adjustment, none of these results were deemed statistically significant.

b) Correlations between After-values for LDL and After-Values for Surface Lipids

In this case, the initial test results suggested that the after-values lipids PC36:02alkyl, PC36:02, LysoPC18:02, PC38:04, PC36:04, PC34:02, PC38:06, PC34:01alkyl, PC34:02alkyl had a statistically significant correlation with the after-values of LDL aggregation. However, after applying the multiple comparisons adjustment, none of these were deemed significant.

3.2. Correlations between LDL Aggregation and Core Lipids

The core lipids were split in the same way as the surface lipids above. The hypotheses were split into a Before-correlations and After-correlations families. The FDR level for Benjamini-Yekutieli adjustment and the p-value level of statistical significance were also as above.

a) Correlations between Before-values for LDL and Before-Values for Core Lipids

None of the correlations between before-values of LDL and before-values of core lipids were deemed statistically significant, both before and after adjusting for multiple comparisons.

b) Correlations between After-values for LDL and After-Values for Core Lipids

The initial test results suggested a statistically significant correlation between the after-values of LDL and the after-values of the lipid ChoE20:04. However, after adjusting for multiple comparisons, none of the results were deemed statistically significant.

4. Are there any significant changes in the clinical parameters of the bariatric patients before and after the operation?

As in the previous question, in this context, "significant changes in clinical parameters" means statistically significant changes in before- and after-values of the clinical parameters. As before, in order to investigate these changes, we formed the differences between the pairs of values for each clinical parameter and visually inspected their normality using histograms. The change in clinical parameters whose differences were normally distributed was tested using the paired t-test. For the ones with non-normal distribution we used the Wilcoxon signed-rank test. The null hypotheses, significance level, and FDR level are the same as in Question 2.
As done previously, we spilt the results into two tables. The first table contains the clinical parameters whose mean difference is statistically significant from 0, evaluated with the paired t-test. The second table contains the clinical parameters whose median of the differences is statistically significant from 0, evaluated with the Wilcoxon signed-rank test.

Paired T-test Results				
Surface Lipid	**Mean Diff.**	**95% CI**	**p-value**	Avg.% Change
RRDiast(Mean)	4.8	(1.86,7.74)	0.0024	0.06
ogtt0	0.67	(0.25,1.09)	0.0029	0.10
ogtt60	2.33	(1.08,3.58)	0.0008	0.26
ogtt120	3.53	(2.34, 4.73)	< 0.00001	0.44
HbA1C	3.21	(0.63, 5.81)	< 0.01682	0.08

Table 5.3: Clinical parameters which have a mean difference of before- and after-operation values that is different from 0 and statistically significant, measured by paired t-test

Wilcoxon Signed-Rank Test Results			
Clinical Parameter	**Median Diff.**	**95% CI**	**p-value**
Weight	25.55	(23.4, 28.7)	< 0.00001
BMI	9.06	(8.24, 10.33)	< 0.000001
Pulse (Mean)	10.00	(5.50, 13.75)	0.00023
ogtt30	-1.75	(-3.35, -0.70)	0.00225

Table 5.4: Clinical parameters which have a median of the differences between before- and after-operation values, which is different from 0 and statistically significant, obtained by Wilcoxon signed-rank test.

5. Are there any significant correlations between LDL aggregation and clinical parameters before and after the operation?

Similarly to the previous questions, in order to investigate the correlations of LDL aggregation and clinical parameters, we split them into before- and after-values. As before, we have two families of hypotheses to test - before and after-correlations, respectively. In order to investigate the correlations, the Spearman's correlation test was used, with p-value < 0.05. The FDR level for the multiple comparisons adjustment used α=0.1.

5.1. Correlations between Before-values for LDL and Before-Values for Clinical Parameters

None of the tests deemed a statistically significant correlation between before-values of LDL and clinical parameters, neither before, nor after adjusting for multiple comparisons.

5.2. Correlations between After-values for LDL and After-Values for Clinical Parameters

In this case, the initial tests deemed the correlation between the after-values of LDL and the after-values of the clinical parameter ogtt0 (Oral glucose tolerance-insulin releasing test value at t=0h) to be statistically significant. However, after applying the multiple comparisons adjustment, none of the results were deemed significant.

6. Are there any significant correlations between change in LDL aggregation, LDL structural lipids and clinical parameters as a result of bariatric surgery?

In order to answer this question, we first formed the differences between pairs of before-after values for LDL, surface lipids, core lipids, and clinical parameters. Once this was done, we then used Spearman's correlation test to evaluate the correlations, with significance level of 0.05. As before, the hypotheses we tested were split into three families - surface-lipids, core-lipids, and clinical-parameters family, respectively. Benjamini-Yekutieli multiple comparisons adjustment with FDR rate α=0.1 was applied to each family.

6.1. Correlations of change in LDL with change in surface lipids

The initial test results yielded a statistically significant correlation between the change in LDL and the change in PC36:02alkyl. However, after applying the Benjamini-Yekutieli multiple comparisons adjustment, none of the results were deemed statistically significant.

6.2. Correlations of change in LDL with change in core lipids

None of the correlations between change in LDL and change in core lipids were deemed statistically significant, neither before, nor after adjusting for multiple comparisons.

6.3. Correlations of change in LDL with change in clinical parameters

None of the correlations between change in LDL and change in clinical parameters were deemed statistically significant, neither before, nor after adjusting for multiple comparisons.

Chapter 6

Discussion

In this thesis we studied the change of LDL aggregation in bariatric patients before and after the surgery. In addition, we studied the changes in the LDL lipidome and clinical parameters of the patients pre- and post operation. We also briefly discussed potential correlations of LDL aggregation with the lipidome and clinical measurements.

We established that there was no statistically significant difference between the before- and after-operation LDL aggregation (where LDL aggregation was described by LDL radius at time t=2h). This result suggests that bariatric surgery does not seem to significantly influence LDL aggregation.

In addition, we established that there were significant changes in surface lipids, both before and after multiple comparisons adjustment. The changes mainly occurred in lipids from the *Lysophosphatidylcholine* (LysoPC), *Phosphatidylcholine* (PC), and *Sphingomyelin* (SPH) groups. As to the core lipids, there were two lipids which exhibited statistically significant difference in before- and after-operation values. However, after adjusting for multiple comparisons, these were deemed insignificant. Thus, bariatric surgery appears to contribute to change in the surface of LDL particles, but does not seem to contribute to change in the core.

Furthermore, we investigated the potential correlations between LDL aggregation and lipids in the surface and core of LDL particles both in before- and after-operation states.
The initial results yielded several statistically significant correlations between LDL aggregation and surface lipids in both before- and after-operation states. However, after these were adjusted for multiple comparisons, none of the correlations were deemed statistically significant.
It is worth pointing out that in the original study by Ruuth et al correlations between LDL aggregation and surface lipids were also studied for two different cohorts. However, none of the initial statistically significant corre-

lations between LDL aggregation and specific surface lipids in our study, in both before- and after-states, coincided with the results from the Ruuth et al study.
As to the core lipids, in the initial results one correlation between a core lipid and LDL aggregation in the after-operation state was found to be significant. However, after multiple comparisons adjustment none of the results were deemed significant. Thus LDL aggregation does not appear to be correlated neither to the surface nor the core structure of the particle in either of the two states.

We then investigated statistically significant changes in the clinical parameters of the patients before and after the operation. Nine out of ten of these clinical factors yielded exhibited statistically significant changes, both before and after multiple comparisons adjustment. Our findings suggested that bariatric surgery appears to contribute to lowering of diastolic blood pressure, pulse, as well as glucose level (in the context of the oral glucose test) in all but one case - at ogtt30, where the glucose level appeared to be higher after the operation. In addition, it appears that a bariatric surgery might contribute towards lowering of the average glucose level, HbA1C. Unsurprisingly, bariatric surgery appears to also lower the weight and BMI of the patients.

Additionally, correlations between LDL aggregation and clinical parameters in both before- and after-operation states were investigated. In the initial results, a statistically significant correlation between LDL and the ogtt0 variable was found in the after-operation state, however, after multiple comparisons adjustment none of the results were deemed significant. Thus, LDL aggregation does not appear to be correlated to the clinical parameters of the patients neither in before-, nor after-operation state.

Lastly, we investigated the correlations between change in LDL aggregation, change in the lipids in the LDL lipidome and clinical parameters. In the initial test results, there was only one surface lipid - PC36:02alkyl, whose change was significantly correlated with the change in LDL aggregation. However, after multiple comparisons adjustment none of the correlations were deemed statistically significant. Thus, bariatric surgery does not seem to induce any simultaneous change of LDL aggregation, lipids from the LDL lipidome, and clinical parameters.

Having discussed the results, it is now worth looking more closely into our methodology and suggest some improvements to be done in the future.

To start with, the LDL aggregation dataset we used to fit our model, was fairly small - 28 subjects, 56 samples - before- and after-operation sample for each subject, 368 points in total. A potential improvement for a future

study would be to obtain a dataset with a larger number of subjects tested. In addition, each sample contained on average around 7 points, representing LDL size at a certain time point. In a future study the number of data points per sample should be increased by measuring the LDL size at smaller time intervals, which would potentially result in creating a more accurate model, decreasing the MAE for predictions.

In addition, we used data points of LDL sizes up to 1500nm, when it is recommended to only use measurements up to 1000nm. If measurements are taken at shorter intervals, this would allow for obtaining a reasonable number of data points with value less than 1000nm. Since the uncertainty around these is considered to be smaller, we can use them to fit a potentially more accurate model. This however raises another question. As mentioned earlier, the upper asymptote for curves fit to aggregation values of LDL, is around 3000nm. One could ask whether using a set of values only up to 1000nm to fit such a curve is reasonable. Thus a logical suggestion would be to look into alternative models to fit to data points up to 1000nm. Since the 1000nm-value for most samples lies at the part of the Richard's curve which is concave upward, one possible suggestion would be to try and model LDL aggregation using exponential growth.

Lastly, while the final model selected for modelling LDL aggregation performs reasonably well in prediction, it could be improved. One way this could be done is by finding more informative prior distributions for the parameters.

Chapter 7

Conclusions

In this thesis we focused on studying LDL aggregation in bariatric surgery patients. The goal of the thesis was to investigate whether LDL aggregation, the structure of LDL, and clinical parameters of patients change significantly after the operation. In addition, we investigated potential correlations of LDL aggregation with lipids in the LDL lipidome and clinical parameters.

The main task in the thesis was to model LDL aggregation and obtain a single quantitative measure for it - in this case, the size of LDL at the 2h time point. After obtaining the predictions for this time point, we conducted statistical tests, in order to answer the research questions about LDL aggregation.

We discovered that undergoing bariatric surgery does not seem to result in significant changes of LDL aggregation in patients. However, it appears to result in changes in the surface of LDL, as well as significant improvement of blood pressure, glucose level and, unsurprisingly, BMI. In addition, LDL aggregation in bariatric surgery patients does not seem to be significantly correlated to the structure of LDL or clinical parameters before or after the operation.
Furthermore, bariatric surgery does not appear to induce any simultaneous changes of LDL aggregation and its structure or clinical parameters of the patients.

This thesis project was based on a fairly small cohort, so replicating the conducted research on data from more subjects could be seen as a logical continuation. Nevertheless, our findings bring us one step closer to revealing the secrets of LDL aggregation and the ultimate goal of making the diagnosis and control of coronary artery disease more reliable and efficient.

Bibliography

[1] NHLBI. High blood cholesterol. webpage. https://www.nhlbi.nih.gov/health-topics/high-blood-cholesterol [Accessed 31 Jul. 2019].

[2] Stan Development Team. Effective sample size. webpage. https://mc-stan.org/docs/2_19/reference-manual/effective-sample-size-section.html [Accessed 31 Jul. 2019].

[3] Anon. Oral glucose tolerance test. webpage. https://www.diabetes.co.uk/oral-glucose-tolerance-test.html [Accessed 31 Jul. 2019].

[4] Anon. What is hba1c? webpage. https://www.diabetes.org.uk/guide-to-diabetes/managing-your-diabetes/hba1c [Accessed 31 Jul. 2019].

[5] Atkinson, A. C. Optimal experimental design. In *International Encyclopedia of the Social Behavioral Sciences*, 2 ed. 2015.

[6] Benjamini, Y., and Yekutieli, D. The control of the false discovery rate in multiple testing under dependency.

[7] Bhagavan, N., and Ha, C. *Essentials of Medical Biochemistry, 2nd Ed.* Academic Press, 2015.

[8] Blanco, A., B. G. *Medical Biochemistry*. Academic Press, 2017, ch. 3.

[9] Burkner, P.-C. brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software 80*, 1 (2017).

[10] Chang, K.-H. *Reliability Analysis*. Elsevier India, 2013, pp. 275–357.

[11] Chavez-Sanchez, L., Chavez-Rueda, K., Legorreta-Haquet, M. V., Montoya-Diaz, E., and Blanco-Favela, F. The innate immune response mediated by tlrs in atherosclerosis. In *Inflammation, Chronic Diseases and Cancer*. IntechOpen, 2012, ch. 3.

[12] DÁvis, T. Wyatt customer support - private correspondence. email, Mar 2019.

[13] DePoy E., L. N. G. *Introduction to Research: Understanding and Applying Multiple Strategies*, 5 ed. Elsevier, 2016.

[14] Fantini, J., and Yahi, N. *Brain lipids in synaptic function and neurological disease.* Academic Press, 2015, pp. 53–85.

[15] Feingold, K. and Grunfeld, C. Introduction to lipids and lipoproteins. ebook, 2018. `https://www.ncbi.nlm.nih.gov/books/NBK305896/` [Accessed 31 Jul. 2019].

[16] Ference, B. e. a. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *European Heart Journal 38*, 32 (2017), 2459–2472.

[17] Gabry, J. Graphical posterior predictive checks using the bayesplot package. website, 2019. `https://cran.r-project.org/web/packages/bayesplot/vignettes/graphical-ppcs.html` [Accessed 2 Aug. 2019].

[18] Gaziano, T., B. A. Growing epidemic of coronary heart disease in low- and middle-income countries. *Current Problems in Cardiology 35*, 2 (2010), 72 – 1150.

[19] Gelman, A. Analysis of variance - why is it more important than ever. *The Annals of Statistics 33*, 1 (2005), 1 – 53.

[20] Gelman, A., Carlin, J. B., Stern, Vehtari, A. D. D. H. S., and Rubin, D. B. *Bayesian Data Analysis*, 3 ed. Chapman and Hall/CRC, 2014.

[21] Gelman, A., J. G. *Handbook of Markov Chain Monte Carlo.*, 1 ed. Chapman Hall/CRC, 2011, ch. 6.

[22] George, S., and Johnson, J. *Atherosclerosis: Molecular and Cellular Mechanisms*. Weinheim: Wiley-VCH-Verl., 2010, ch. 1.3.

[23] Goeman J. J., S. A. Tutorial in biostatistics: multiple hypothesis testing in genomics. *Statistics in Medicine 33*, 11 (2014), 1946–1978.

[24] Graphpad. Interpreting results: Wilcoxon signed-rank test. webpage. `https://www.graphpad.com/guides/prism/7/statistics/stat_interpreting_results_wilcoxon_.htm?toc=0&printWindow` [Accessed 31 Jul. 2019].

[25] Grundy, S. M. Obesity, metabolic syndrome, and cardiovascular disease. *The Journal of Clinical Endocrinology Metabolism, 89*, 6, 2595â2600.

[26] J., F. *Applied Regression Analysis and Generalized Linear Models*. Sage, 2015, ch. 15.

[27] Kahilakoski, O.-P. Bayesian regression analysis of sickness absence. Master's thesis, Aalto University School of Science and Technology, 2011.

[28] Maymon, G. *Stochastic Crack Propagation: Essential Practical Aspects*. Elsevier, ch. 2, pp. 9–18.

[29] McDonald, J. *Handbook of Biological Statistics*, 3 ed. Sparky House Publishing, 2014.

[30] NCBI. What is blood pressure and how is it measured? webpage. `https://www.ncbi.nlm.nih.gov/books/NBK279251/` [Accessed 31 Jul. 2019].

[31] NHLBI. Atherosclerosis. webpage. `https://www.nhlbi.nih.gov/health-topics/atherosclerosis` [Accessed 31 Jul. 2019].

[32] NIDDK. Definition facts for bariatric surgery. webpage. `https://www.niddk.nih.gov/health-information/weight-management/bariatric-surgery/definition-facts` [Accessed 2 Aug. 2019].

[33] Paar, Anton. The principles of dynamic light scattering :: Anton paar wiki. webpage. `https://wiki.anton-paar.com/en/the-principles-of-dynamic-light-scattering/` [Accessed 31 Jul. 2019].

[34] Richards, F. A flexible growth function for empirical use. *Journal of Experimental Botany 10*, 2, 290–301.

[35] Ruuth, M. e. a. Susceptibility of low-density lipoprotein particles to aggregate depends on particle lipidome, is modifiable, and associates with future cardiovascular deaths. *European Heart Journal 39*, 27 (2018), 2562–2573.

[36] Stan Development Team. Interfaces. forum. `https://discourse.mc-stan.org/c/interfaces` [Accessed 31 Jul. 2019].

[37] Suckow, M. *The Laboratory Rabbit, Guinea Pig, Hamster, and Other Rodents.* Elsevier, 2012.

[38] Tampere University of Technology. Bayesian estimation. webpage. http://www.cs.tut.fi/~hehu/SSP/lecture10.pdf [Accessed 2 Aug. 2019].

[39] Theodoridis, S. *Machine Learning.* O'Reily, 2015, ch. 2.

[40] Vehtari A., Gelman, A. J. G. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing 27*, 5 (2017), 1413–1432.

[41] Weisenstein, E. Pearson's skewness coefficients. webpage. http://mathworld.wolfram.com/PearsonsSkewnessCoefficients.html [Accessed 31 Jul. 2019].

[42] Wyatt-Technology. Dynamics user's guide. webpage. https://physiology.case.edu/media/eq_manuals/eq_manual_Dynamics_Users_Guide_M1400_Rev_K.pdf [Accessed 31 Jul. 2019].

[43] Wyatt.com. Dynamic light scattering (dls) and zeta potential detectors. webpage. https://www.wyatt.com/products/instruments/dynamic-electrophoretic-light-scattering-detectors.html [Accessed 31 Jul. 2019].

[44] Wyatt.com. The theory behind dynamic light scattering (dls). webpage. https://www.wyatt.com/library/theory/dynamic-light-scattering-theory.html [Accessed 31 Jul. 2019].

Appendix A

Appendix

A.1 Model Output

A.1.1 Output from Model 1

```
summary(model1)
 Family: gamma
  Links: mu = log; shape = identity
Formula: LDL ~ log(15 + (alpha - 15)/(1 + exp((gamma - Time)/delta)))
         alpha ~ 1 + Cat + (1 | Pat)
         gamma ~ 1 + Cat + (1 | Pat)
         delta ~ 1
   Data: Data (Number of observations: 368)
Samples: 4 chains, each with iter = 5000; warmup = 1000; thin = 1;
         total post-warmup samples = 16000

Group-Level Effects:
~Pat (Number of levels: 28)
                    Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
sd(alpha_Intercept)   758.03    204.21   439.35  1230.83       5936 1.00
sd(gamma_Intercept)     0.16      0.05     0.08     0.26       3461 1.00

Population-Level Effects:
                Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
alpha_Intercept  1537.57    257.87  1105.89  2111.64       6459 1.00
alpha_CatB        174.07    144.22     4.27   514.51       4008 1.00
gamma_Intercept     2.37      0.06     2.25     2.50       8699 1.00
gamma_CatB          0.08      0.04     0.01     0.18       4168 1.00
delta_Intercept     0.27      0.01     0.25     0.29      12127 1.00

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
shape     6.81      0.53     5.81     7.89      15535 1.00

Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample
is a crude measure of effective sample size, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

A.1.2 Output from Model 2

```
Family: gamma
  Links: mu = log; shape = identity
Formula: LDL ~ log(15 + (alpha - 15)/(1 + exp((gamma - Time)/delta)))
         alpha ~ 1 + (1 | Pat/Cat)
         gamma ~ 1 + (1 | Pat/Cat)
         delta ~ 1
   Data: Data (Number of observations: 368)
Samples: 4 chains, each with iter = 5000; warmup = 1000; thin = 1;
         total post-warmup samples = 16000

Group-Level Effects:
~Pat (Number of levels: 28)
                     Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
sd(alpha_Intercept)  1241.56    389.34   389.35  2004.24        2840 1.00
sd(gamma_Intercept)     0.05      0.03     0.00     0.12        3956 1.00

~Pat:Cat (Number of levels: 56)
                     Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
sd(alpha_Intercept)  1225.42    330.18   674.34  1945.87        2378 1.00
sd(gamma_Intercept)     0.07      0.03     0.01     0.12        2195 1.00

Population-Level Effects:
               Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
alpha_Intercept  2982.67    458.36  2180.91  3972.90        8102 1.00
gamma_Intercept     2.64      0.05     2.54     2.74        9394 1.00
delta_Intercept     0.28      0.01     0.27     0.30       13005 1.00

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
shape    13.74      1.10    11.67    15.95       19207 1.00

Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample
is a crude measure of effective sample size, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

A.2 Pareto-k Diagnostic Plots

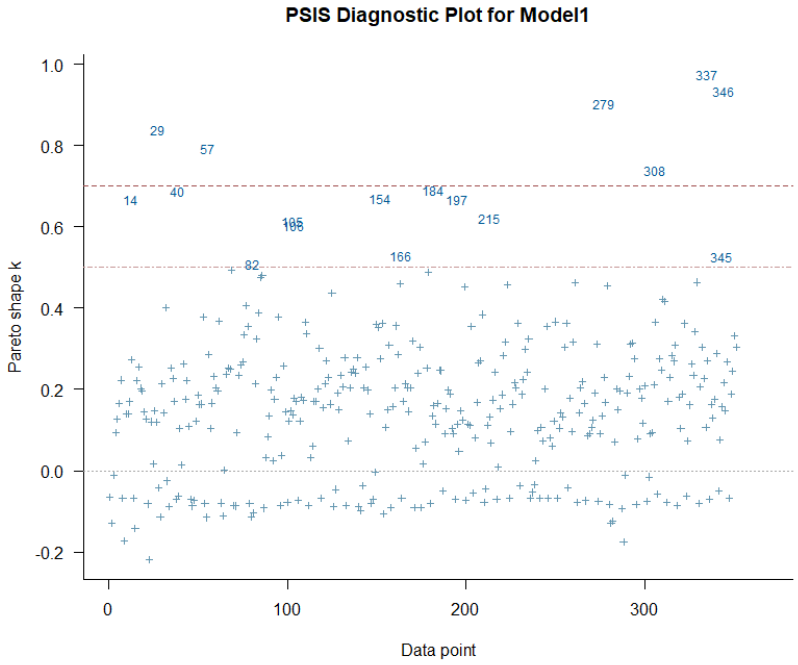

Figure A.1: Pareto-k Diagnostic Plot for Model 1

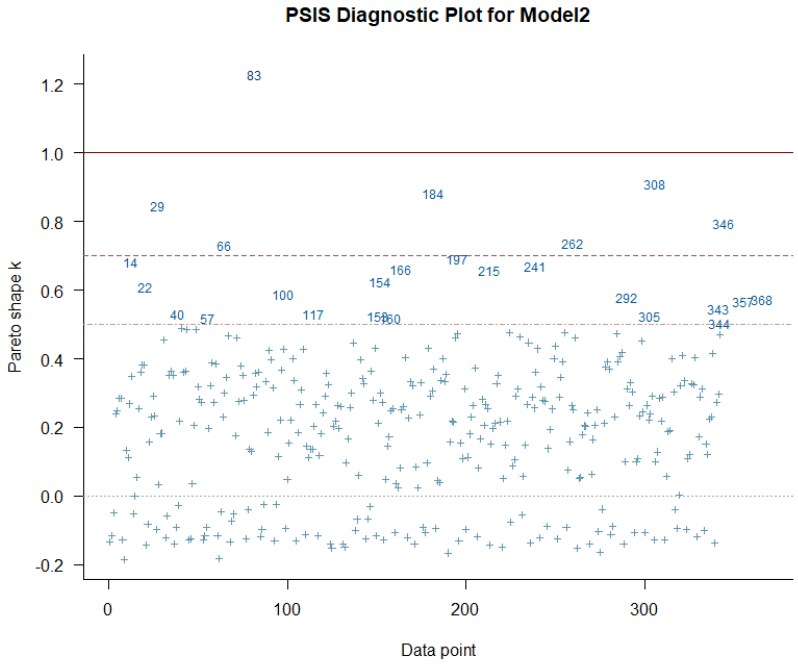

Figure A.2: Pareto-k Diagnostic Plot for Model 2